

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 05:37:18 ; Search time 3312.9 Seconds
(without alignments)
132.650 Million cell updates/sec

Title: US-09-235-416-3
Perfect score: 21
Sequence: 1 atgtcggcggtggaataatc 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query
No. Score Match Length DB ID Description

1	18.4	87.6	143981	9	HSJ1050K3	AL121776 Human DNA
c	17.8	84.8	204067	2	AL662786	AL662786 Mus muscu
c	16.8	80.0	737	2	HS336359	AJ336359 Homo sapi
c	16.8	80.0	750	9	HS336191	AJ336191 Homo sapi
c	16.8	80.0	792	9	HS334149	AJ334149 Homo sapi
c	16.8	80.0	88941	2	AC106623	AC106623 Rattus no
c	16.8	80.0	100912	2	AC021089	AC021089 Homo sapi
c	16.8	80.0	104188	2	AF004257	AF004257 Oryza sat
c	16.8	80.0	107025	2	AL139235_3	Continuation (4 of
c	16.8	80.0	107200	2	AC106570	AC106570 Rattus no
c	16.8	80.0	194322	3	AC024791	AC024791 Caenorhab
c	16.8	80.0	319857	2	AC006782	AC006782 Caenorhab
c	16.8	80.0	57013	2	AC015769	AC015769 Homo sapi
c	16.8	80.0	63528	2	AC091196	AC091196 Homo sapi
c	16.8	80.0	100511	2	AC010774	AC010774 Homo sapi
c	16.8	80.0	156929	2	AC073751	AC073751 Mus muscu
c	16.8	80.0	196057	2	AC073725	AC073725 Mus muscu
c	16.2	77.1	3505	8	AB045127	AB045127 Schizosac
c	16.2	77.1	27777	3	AC002447	AC002447 Drosophill
c	16.2	77.1	42844	3	CEC35D6	282261 Caenorhabd
c	16.2	77.1	44751	8	SPCC16C4	AL031535 S.pombe c
c	16.2	77.1	56780	2	AC107297	AC107297 Homo sapi
c	16.2	77.1	58950	2	AC102353	AC102353 Mus muscu
c	16.2	77.1	91202	2	AC007609	AC007609 Homo sapi
c	16.2	77.1	110000	2	LMFLCHR18_04	Continuation (5 of
c	16.2	77.1	119558	2	AC097782	AC097782 Rattus no
c	16.2	77.1	138877	2	AF004398	AF004398 Oryza sat
c	16.2	77.1	140688	2	AP003801	AP003801 Oryza sat
c	16.2	77.1	148366	2	AC108717	AC108717 Homo sapi
c	16.2	77.1	151345	2	AP003332	AP003332 Oryza sat
c	16.2	77.1	152147	9	AC091842	AC091842 Homo sapi
c	16.2	77.1	157774	2	AC013353	AC013353 Trypanoso
c	16.2	77.1	163954	2	AC017155	AC017155 Drosophill
c	16.2	77.1	164558	2	AC040959	AC040959 Homo sapi
c	16.2	77.1	165507	2	AC023823	AC023823 Homo sapi
c	16.2	77.1	166941	2	AC093994	AC093994 Rattus no
c	16.2	77.1	175223	9	AC004617	AC004617 Homo sapi
c	16.2	77.1	177863	2	AL359388	AL359388 Homo sapi
c	16.2	77.1	178454	2	AC055776	AC055776 Mus muscu
c	16.2	77.1	180401	9	AC007489	AC007489 Homo sapi
c	16.2	77.1	185063	2	AC104036	AC104036 Homo sapi
c	16.2	77.1	186432	3	AC099034	AC099034 Drosophill
c	16.2	77.1	190803	9	AL355490	AL355490 Human DNA
c	16.2	77.1	302422	3	AE003455	AE003455 Drosophill
c	16.2	77.1	303040	1	AP000991	AP000991 Thermopia

ALIGNMENTS

RESULT 1
HSJ1050K3

LOCUS
DEFINITION

143981 bp DNA linear PRI 04-APR-2001
Human DNA sequence from clone RP5-1050K3 on chromosome 20. Contains part of the EYA2 gene for eyes absent (Drosophila) homolog 2, a Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) pseudogene, an RPL27A (60S ribosomal protein L27A) pseudogene, ESTs, STSs and GSSs, complete sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

AL121776
AL121776.19 GI-8218094
HTG; EYA2; GAPDH; RPL27A.
human.

ORGANISM

Homo sapiens

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 143981)
Peck,A.

TITLE
JOURNAL

Direct Submission
Submitted (03-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

COMMENT

On Jun 3, 2000 this sequence version replaced gi:7799934.
During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr20>
This sequence is the entire insert of clone RP5-1050K3. The true right end of clone RP11-323C15 is at 58476 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-1050K3 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>
VECTOR: pCYPAC2.

FEATURES

source	Location/Qualifiers
	1. .143981
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="20"
	/clone="RP5-1050K3"
	/clone_lib="RPCI-5"
	134. .175
repeat_region	/note="L2 repeat: matches 2709. .2750 of consensus"
repeat_region	485. .664
repeat_region	/note="MIR repeat: matches 46. .260 of consensus"
repeat_region	774. .969
repeat_region	/note="MIR repeat: matches 70. .262 of consensus"
repeat_region	1091. .1330
repeat_region	/note="MER46A repeat: matches 2. .235 of consensus"
repeat_region	1387. .1679
repeat_region	/note="AluX repeat: matches 1. .294 of consensus"
repeat_region	2480. .2676
misc_feature	/note="MIR repeat: matches 41. .262 of consensus"
repeat_region	complement(2633. .2991)
repeat_region	/note="match: GSS: Em:AQ017978"
repeat_region	2847. .2874
misc_feature	/note="14 copies 2 mer tt 89% conserved"
misc_feature	2926. .3357
misc_feature	/note="match: GSS: Em:AQ423175"
repeat_region	2930. .3460
repeat_region	/note="match: GSS: Em:AQ423187"
repeat_region	3022. .3103
repeat_region	/note="L2 repeat: matches 2669. .2750 of consensus"
repeat_region	3109. .3326
repeat_region	/note="MIR repeat: matches 20. .246 of consensus"
repeat_region	3344. .3543
repeat_region	/note="LIMC4 repeat: matches 7750. .7977 of consensus"
repeat_region	3544. .4059
repeat_region	/note="MUT1F repeat: matches 1. .541 of consensus"
repeat_region	4060. .4192
repeat_region	/note="LIMC4 repeat: matches 7626. .7750 of consensus"
repeat_region	5002. .5192
misc_feature	/note="MIR repeat: matches 52. .248 of consensus"
repeat_region	complement(5198. .5634)
repeat_region	/note="match: GSS: Em:AQ787200"
repeat_region	5411. .5684
repeat_region	/note="MIR repeat: matches 1. .262 of consensus"
misc_feature	5646. .6090
repeat_region	/note="match: GSS: Em:AQ609792"
repeat_region	6373. .6495
repeat_region	/note="MIR repeat: matches 53. .192 of consensus"
repeat_region	6517. .6723
repeat_region	/note="MIR repeat: matches 4. .215 of consensus"
repeat_region	6768. .6950
repeat_region	/note="MIR repeat: matches 63. .245 of consensus"
repeat_region	6951. .7253
repeat_region	/note="AluJo repeat: matches 1. .305 of consensus"
repeat_region	7254. .7265
repeat_region	/note="MIR repeat: matches 245. .256 of consensus"
repeat_region	7395. .7610
repeat_region	/note="AluJo repeat: matches 95. .310 of consensus"
repeat_region	7809. .7952
repeat_region	/note="MIR repeat: matches 28. .184 of consensus"
repeat_region	7933. .8262
repeat_region	/note="AluX repeat: matches 1. .312 of consensus"
repeat_region	8263. .8288
repeat_region	/note="MIR repeat: matches 184. .207 of consensus"
repeat_region	8289. .8450
repeat_region	/note="MER20 repeat: matches 1. .164 of consensus"
repeat_region	8596. .8676
repeat_region	/note="MIR repeat: matches 109. .199 of consensus"
repeat_region	9736. .9877
repeat_region	/note="MIR repeat: matches 109. .255 of consensus"
repeat_region	10029. .10325
repeat_region	/note="AluJB repeat: matches 11. .295 of consensus"
repeat_region	10370. .10506
repeat_region	/note="MIR repeat: matches 103. .241 of consensus"
repeat_region	11776. .11966
repeat_region	/note="MER58A repeat: matches 6. .224 of consensus"
repeat_region	12110. .12326
repeat_region	/note="MER58A repeat: matches 4. .223 of consensus"
repeat_region	12716. .12835
repeat_region	/note="MLT1I repeat: matches 286. .407 of consensus"
repeat_region	13032. .13336
repeat_region	/note="AluJo repeat: matches 1. .312 of consensus"
repeat_region	13374. .13594
repeat_region	/note="MER58A repeat: matches 4. .224 of consensus"
repeat_region	13690. .13795
repeat_region	/note="L2 repeat: matches 2583. .2705 of consensus"
misc_feature	14045. .14421
repeat_region	/note="match: GSS: Em:AQ031759"
repeat_region	14143. .14337
repeat_region	/note="MIR repeat: matches 1. .204 of consensus"
misc_feature	complement(14410. .14921)
repeat_region	/note="match: GSS: Em:AQ055749"
repeat_region	14664. .14692
repeat_region	/note="L2 repeat: matches 164. .192 of consensus"
repeat_region	14853. .14895
repeat_region	/note="L2 repeat: matches 2707. .2749 of consensus"
repeat_region	14981. .15113
repeat_region	/note="MIR repeat: matches 4. .144 of consensus"
repeat_region	15094. .15249
repeat_region	/note="MIR repeat: matches 9. .177 of consensus"
repeat_region	15227. .15297
repeat_region	/note="L2 repeat: matches 2621. .2689 of consensus"
misc_feature	complement(16294. .16547)
misc_feature	/note="match: GSS: Em:B14853"
misc_feature	17097. .17504
repeat_region	/note="match: GSS: Em:AQ128898"
repeat_region	17149. .17400
repeat_region	/note="L2 repeat: matches 2468. .2740 of consensus"
repeat_region	17433. .17605
repeat_region	/note="L2 repeat: matches 2192. .2374 of consensus"
repeat_region	17676. .17729
repeat_region	/note="L2 repeat: matches 2695. .2750 of consensus"
repeat_region	18336. .18486
repeat_region	/note="MIR repeat: matches 98. .261 of consensus"
repeat_region	18770. .18941
repeat_region	/note="MER20 repeat: matches 3. .184 of consensus"
repeat_region	19066. .19554
repeat_region	/note="L2 repeat: matches 2157. .2709 of consensus"
repeat_region	20336. .20489
repeat_region	/note="AluSp repeat: matches 140. .298 of consensus"

```
repeat_region 21031..21255
/note="MER58A repeat: matches 1..224 of consensus"
repeat_region 21526..21609
/note="MLT11 repeat: matches 322..410 of consensus"
repeat_region 21723..21925
/note="MLT1J repeat: matches 19..261 of consensus"
repeat_region 21942..22258
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region 22267..22337
/note="MIR repeat: matches 180..250 of consensus"
repeat_region 22535..23043
/note="MLT1D repeat: matches 7..497 of consensus"
misc_feature 23092..23310
/note="match: GSS: Em:AQ306389"
repeat_region 23762..24321
/note="L1PA13 repeat: matches 5573..6142 of consensus"
misc_feature complement(24352..24917)
/note="match: GSS: Em:AQ667419"
repeat_region 24486..24656
/note="MER20 repeat: matches 31..218 of consensus"
repeat_region 24804..24865
/note="31 copies 2 mer ca 72% conserved"
repeat_region 25351..25559
/note="MIR repeat: matches 32..249 of consensus"
repeat_region 25564..25754
/note="MIR repeat: matches 77..262 of consensus"
repeat_region 25803..26008
/note="MER58A repeat: matches 9..224 of consensus"
repeat_region 26250..26541
/note="AluSk repeat: matches 1..301 of consensus"
repeat_region 26544..27216
/note="L1MA5 repeat: matches 5469..6149 of consensus"
repeat_region 27217..27523
87.6%; Score 18.4; DB 9; Length 143981;
Query Match Best Local Similarity 95.0%; Pred No.1.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaatat 20
||||| |||||||
Db 82141 ATGTGGGTGGTGGAAATAT 82160

RESULT 2
AL662786/c 204067 bp DNA linear HTG 04-JAN-2002
LOCUS Mus musculus chromosome 11 clone RP23-340010, *** SEQUENCING IN
DEFINITION PROGRESS ***, in unordered pieces.
ACCESSION AL662786
VERSION AL662786.2 GI:17903997
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
McLay, K.
Direct Submission
Submitted (03-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Dec 18, 2001 this sequence version replaced gi:17736652.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bm340010
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
```

```
Consensus quality: 202716 bases at least Q40
Consensus quality: 203013 bases at least Q30
Consensus quality: 203237 bases at least Q20
Insert size: 203467; sum-of-contigs
Insert size: 198131; 4.7% error; agarose-fp
Quality coverage: 10.99x in Q20 bases; sum-of-contigs Quality
coverage: 11.59x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

FEATURES

```
source .
1..204067
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-340010"
/clone_lib="RPC1-23"
1..9424
/note="assembly_fragment:02278
fragment_chain:1
clone_end:SP6
vector_side:left"
misc_feature 9525..31155
/note="assembly_fragment:00677
fragment_chain:1"
misc_feature 31256..89487
/note="assembly_fragment:01888
fragment_chain:1"
misc_feature 89588..119319
/note="assembly_fragment:04012
fragment_chain:1"
misc_feature 119420..175677
/note="assembly_fragment:03857
fragment_chain:1"
misc_feature 175778..197522
/note="assembly_fragment:02822
fragment_chain:1"
misc_feature 197623..204067
/note="assembly_fragment:03165
fragment_chain:1
clone_end:Y7
vector_side:right"
BASE COUNT 57719 a 40407 c 41655 g 63676 t 610 others
ORIGIN
```

Query Match 84.8%; Score 17.8; DB 2; Length 204067;
Best Local Similarity 90.5%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaatatc 21
||||| |||||||

Db 94801 ATGTGGGAGGTGGAAATATC 94781

RESULT 3

```
HSA336359/c 737 bp DNA linear PRI 01-OCT-2001
LOCUS HSA336359
DEFINITION Homo sapiens genomic sequence surrounding NotI site, clone
HSJ-CB3RS.
ACCESSION AJ336359
VERSION AJ336359.1 GI:15880777
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 737)
REFERENCE Kutsenko, A. S., Gizatullin, R. Z., Al-Amin, A. N., Wang, F.,
Podowski, R. M., Matushkin, Y. G., Kvasha, S. M., Gyanchandani, A.,
Muravenko, O. V., Protopopov, A. I., Kashuba, V. I., Kisselev, L. L.,
```

RESULT 5

Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
 Lounsbury, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,
 Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
 Ogihara, M., Okunishi, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Peir, J., Perez, J., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M.,
 Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoshitaishvili, N.,
 Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 88941)
 Worley, K.C.

Direct Submission
 Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GLSB
 Center clone name: CH230-224B2
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to
 findPhrapList
 Consensus quality: 64203 bases at least Q40
 Consensus quality: 70064 bases at least Q30
 Consensus quality: 75160 bases at least Q20
 Estimated insert size: 56201; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 0.6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 46 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 3507: contig of 3507 bp in length
 * 3508 3607: gap of unknown length
 * 3608 7003: contig of 3396 bp in length
 * 7004 7103: gap of unknown length
 * 7104 10062: contig of 2959 bp in length
 * 10063 10162: gap of unknown length
 * 10163 13524: contig of 3362 bp in length
 * 13525 16718: gap of unknown length
 * 13625 16718: contig of 3094 bp in length
 * 16719 16818: gap of unknown length
 * 16819 20154: contig of 3336 bp in length
 * 20155 20254: gap of unknown length
 * 20255 22778: contig of 2524 bp in length
 * 22779 22878: gap of unknown length

22879 25169: contig of 2291 bp in length
 25170 25269: gap of unknown length
 25270 27417: contig of 2148 bp in length
 27418 27517: gap of unknown length
 27519 29501: contig of 1983 bp in length
 29502 29601: gap of unknown length
 29602 31479: contig of 1879 bp in length
 31480 31579: gap of unknown length
 31580 34809: contig of 3230 bp in length
 34810 34909: gap of unknown length
 34910 36936: contig of 2027 bp in length
 36937 37036: gap of unknown length
 37037 38629: contig of 1593 bp in length
 38630 38729: gap of unknown length
 38730 40371: contig of 1642 bp in length
 40372 40471: gap of unknown length
 40472 41983: contig of 1512 bp in length
 41984 42083: gap of unknown length
 42084 44236: contig of 2153 bp in length
 44237 44336: gap of unknown length
 44337 46111: contig of 1775 bp in length
 46112 46211: gap of unknown length
 46212 47884: contig of 1673 bp in length
 47885 47984: gap of unknown length
 47986 50053: contig of 2069 bp in length
 50054 50153: gap of unknown length
 50154 51358: contig of 1205 bp in length
 51359 51458: gap of unknown length
 51459 53537: contig of 2079 bp in length
 53538 53637: gap of unknown length
 53639 55158: contig of 1421 bp in length
 55159 56396: contig of 1238 bp in length
 56397 56496: gap of unknown length
 56497 57898: contig of 1402 bp in length
 57899 57998: gap of unknown length
 57999 59678: contig of 1880 bp in length
 59679 59778: gap of unknown length
 59779 60801: contig of 1022 bp in length
 60802 61958: gap of unknown length
 61959 62058: contig of 1058 bp in length
 62059 63787: gap of unknown length
 63788 63887: gap of unknown length
 63889 65032: contig of 1145 bp in length
 65040 65132: gap of unknown length
 65133 67086: contig of 1954 bp in length
 67087 67186: gap of unknown length
 67187 69261: contig of 2075 bp in length
 69262 69361: gap of unknown length
 69362 70708: contig of 1347 bp in length
 70709 70808: gap of unknown length
 70809 71876: contig of 1068 bp in length
 71877 71976: gap of unknown length
 71977 73010: contig of 1034 bp in length
 73011 73110: gap of unknown length
 73111 74254: contig of 1144 bp in length
 74255 74354: gap of unknown length
 74355 75787: contig of 1433 bp in length
 75788 75887: gap of unknown length
 75888 77293: contig of 1406 bp in length
 77294 77393: gap of unknown length
 77394 78438: contig of 1045 bp in length
 78439 78538: gap of unknown length
 78539 79665: contig of 1127 bp in length
 79666 79765: gap of unknown length
 79766 81448: contig of 1683 bp in length
 81449 81548: gap of unknown length
 81549 82990: contig of 1342 bp in length
 82991 84611: contig of 1621 bp in length
 84612 84711: gap of unknown length
 84712 86141: contig of 1430 bp in length

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

<p>* 86142 86241: gap of unknown length</p> <p>* 86242 87695: contig of 1458 bp in length</p> <p>* 87700 87795: gap of unknown length</p> <p>* 87800 88941: contig of 1142 bp in length.</p> <p>Location/Qualifiers</p> <p>1. .88941</p> <p>/organism="Rattus norvegicus"</p> <p>/db_xref="taxon:10116"</p> <p>/clone="CH230-224B2"</p> <p>23763 a 19418 c 16127 g 25100 t 4533 others</p> <p>BASE COUNT</p> <p>ORIGIN</p> <p>Query Match 80.0%; Score 16.8; DB 2; Length 88941;</p> <p>Best Local Similarity 90.0%; Pred. No. 8.1e+02;</p> <p>Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;</p> <p>QY 2 tgtcggcggtggaataatc 21</p> <p> </p> <p>Db 19329 TGTGGGTGTGGAATCTC 19310</p> <p>RESULT 7</p> <p>AC021089/c</p> <p>DEFINITION</p> <p>AC021089</p> <p>VERSION</p> <p>KEYWORDS</p> <p>SOURCE</p> <p>ORGANISM</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>TITLE</p> <p>JOURNAL</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>TITLE</p> <p>JOURNAL</p> <p>COMMENT</p>	<p>21140: gap of unknown length</p> <p>21041 24617: contig of 3477 bp in length</p> <p>21141 24618 24717: gap of unknown length</p> <p>24718 27291: contig of 2574 bp in length</p> <p>27292 27391: gap of unknown length</p> <p>36976 36976: contig of 9585 bp in length</p> <p>37076 37076: gap of unknown length</p> <p>47599 47599: contig of 10523 bp in length</p> <p>47600 47699: gap of unknown length</p> <p>47700 68271: contig of 20572 bp in length</p> <p>68272 80965: contig of 12594 bp in length</p> <p>80966 81065: gap of unknown length</p> <p>81066 87307: contig of 6242 bp in length</p> <p>87308 87407: gap of unknown length</p> <p>87408 97701: contig of 10294 bp in length</p> <p>97702 97801: gap of unknown length</p> <p>97802 100912: contig of 3111 bp in length.</p> <p>Location/Qualifiers</p> <p>1. .100912</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/chromosome="16"</p> <p>/clone="CTD-2264D9"</p> <p>/clone_lib="Caltech human BAC library D"</p> <p>27774 a 20918 c 22009 g 29159 t 1052 others</p> <p>BASE COUNT</p> <p>ORIGIN</p> <p>Query Match 80.0%; Score 16.8; DB 2; Length 100912;</p> <p>Best Local Similarity 90.0%; Pred. No. 8.1e+02;</p> <p>Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;</p> <p>QY 2 tgtcggcggtggaataatc 21</p> <p> </p> <p>Db 84084 TCTCGGCTGTGGAATATC 84065</p> <p>RESULT 8</p> <p>AP004257/c</p> <p>LOCUS</p> <p>DEFINITION</p> <p>PROGRESS ***; in ordered pieces.</p> <p>AP004257</p> <p>VERSION</p> <p>KEYWORDS</p> <p>SOURCE</p> <p>ORGANISM</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>TITLE</p> <p>JOURNAL</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>TITLE</p> <p>JOURNAL</p> <p>COMMENT</p>	<p>86241: gap of unknown length</p> <p>87695: contig of 1458 bp in length</p> <p>87795: gap of unknown length</p> <p>88941: contig of 1142 bp in length.</p> <p>Location/Qualifiers</p> <p>1. .88941</p> <p>/organism="Rattus norvegicus"</p> <p>/db_xref="taxon:10116"</p> <p>/clone="CH230-224B2"</p> <p>23763 a 19418 c 16127 g 25100 t 4533 others</p> <p>BASE COUNT</p> <p>ORIGIN</p> <p>Query Match 80.0%; Score 16.8; DB 2; Length 88941;</p> <p>Best Local Similarity 90.0%; Pred. No. 8.1e+02;</p> <p>Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;</p> <p>QY 2 tgtcggcggtggaataatc 21</p> <p> </p> <p>Db 19329 TGTGGGTGTGGAATCTC 19310</p> <p>RESULT 7</p> <p>AC021089/c</p> <p>DEFINITION</p> <p>AC021089</p> <p>VERSION</p> <p>KEYWORDS</p> <p>SOURCE</p> <p>ORGANISM</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>TITLE</p> <p>JOURNAL</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>TITLE</p> <p>JOURNAL</p> <p>COMMENT</p>	<p>21140: gap of unknown length</p> <p>21041 24617: contig of 3477 bp in length</p> <p>21141 24618 24717: gap of unknown length</p> <p>24718 27291: contig of 2574 bp in length</p> <p>27292 27391: gap of unknown length</p> <p>36976 36976: contig of 9585 bp in length</p> <p>37076 37076: gap of unknown length</p> <p>47599 47599: contig of 10523 bp in length</p> <p>47600 47699: gap of unknown length</p> <p>47700 68271: contig of 20572 bp in length</p> <p>68272 80965: contig of 12594 bp in length</p> <p>80966 81065: gap of unknown length</p> <p>81066 87307: contig of 6242 bp in length</p> <p>87308 87407: gap of unknown length</p> <p>87408 97701: contig of 10294 bp in length</p> <p>97702 97801: gap of unknown length</p> <p>97802 100912: contig of 3111 bp in length.</p> <p>Location/Qualifiers</p> <p>1. .100912</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/chromosome="16"</p> <p>/clone="CTD-2264D9"</p> <p>/clone_lib="Caltech human BAC library D"</p> <p>27774 a 20918 c 22009 g 29159 t 1052 others</p> <p>BASE COUNT</p> <p>ORIGIN</p> <p>Query Match 80.0%; Score 16.8; DB 2; Length 100912;</p> <p>Best Local Similarity 90.0%; Pred. No. 8.1e+02;</p> <p>Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;</p> <p>QY 2 tgtcggcggtggaataatc 21</p> <p> </p> <p>Db 84084 TCTCGGCTGTGGAATATC 84065</p> <p>RESULT 8</p> <p>AP004257/c</p> <p>LOCUS</p> <p>DEFINITION</p> <p>PROGRESS ***; in ordered pieces.</p> <p>AP004257</p> <p>VERSION</p> <p>KEYWORDS</p> <p>SOURCE</p> <p>ORGANISM</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>TITLE</p> <p>JOURNAL</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>TITLE</p> <p>JOURNAL</p> <p>COMMENT</p>
---	---	---	---

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES

source
 1. 104188
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="2"
 /clone="QJ1369.G08"
 30752 a 21170 c 21376 g 30890 t

BASE COUNT

ORIGIN

Query Match 80.0%; Score 16.8; DB 2; Length 104188;
 Best Local Similarity 90.0%; Pred. No. 8.1e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaatat 20
 ||||| ||||| ||||| |||||
 Db 54563 ATGTCGGCGGCAGAAATAT 54544

RESULT 9

AL139235_3

WPCOMMENT

Sequence split into 4 fragments LOCUS AL139235 Accession AL139235

Fragment Name	Begin	End
AL139235_0	1	110000
AL139235_1	100001	210000
AL139235_2	200001	310000
AL139235_3	300001	410025

Continuation (4 of 4) of AL139235 from base 300001 (AL139235 Homo sapiens chromosome 9 c

Query Match

Best Local Similarity 80.0%; Score 16.8; DB 2; Length 107025;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgtcggcggtggaatatc 21
 ||||| ||||| ||||| |||||

Db 16779 TCTCGGCGTGTGAATATC 16798

RESULT 10

AC106570/c

LOCUS

AC106570 172300 bp DNA linear HTG 12-JAN-2002
 Rattus norvegicus clone CH230-119J12, *** SEQUENCING IN PROGRESS
 *** 67 unordered pieces.

ACCESSION

AC106570

VERSION

AC106570.1 GI:18139094

KEYWORDS

HTG; HTGS_PHASE1.

SOURCE

Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 172300)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbaria,J.,

Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,F., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
 Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Lounsged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
 Ogdu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Pery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,K., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
 Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
 Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wlaczky,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 172300)

Worley,K.C.

Direct Submission

JOURNAL

TITLE

JOURNAL

COMMENT

JOURNAL

TITLE

JOURNAL

COMMENT

JOURNAL

TITLE

JOURNAL

COMMENT

JOURNAL

TITLE

JOURNAL

COMMENT

JOURNAL

TITLE

JOURNAL

COMMENT

JOURNAL

TITLE

JOURNAL

COMMENT

JOURNAL

TITLE

JOURNAL

COMMENT

JOURNAL

TITLE

JOURNAL

COMMENT

JOURNAL

TITLE

JOURNAL

COMMENT

JOURNAL

TITLE

JOURNAL

COMMENT

JOURNAL

TITLE

JOURNAL

COMMENT

JOURNAL

TITLE

JOURNAL

COMMENT

Assembly program: Phrap; version 0.990329First call to

findPhrapList

Consensus quality: 135596 bases at least Q40

Consensus quality: 143040 bases at least Q30

Consensus quality: 148817 bases at least Q20

Estimated insert size: 136955; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 67 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 9083: contig of 9083 bp in length

* 9084 9183: gap of unknown length

* 9184 17988: contig of 8805 bp in length

* 17989 18088: gap of unknown length

* 18089 22682: contig of 4594 bp in length

* 22683 22782: gap of unknown length

* 22783 27986: contig of 5204 bp in length

* 27987 28086: gap of unknown length

* 28087 31496: contig of 3410 bp in length

* 31497 31596: gap of unknown length

* 31597 36576: contig of 4980 bp in length

* 36577 41576: gap of unknown length

* 41576 41576: contig of 4900 bp in length

REFERENCE
AUTHORS
TITLE
JOURNAL

6 (bases 1 to 194322)
Waterston,R.
Direct Submission
Submitted (06-APR-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

7 (bases 1 to 194322)
Waterston,R.
Direct Submission
Submitted (30-AUG-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

8 (bases 1 to 194322)
Waterston,R.
Direct Submission
Submitted (05-OCT-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

9 (bases 1 to 194322)
Waterston,R.
Direct Submission
Submitted (16-NOV-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

COMMENT

Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis , MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: rw@nematoe.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one ml3 subclone.

NEIGHBORING COSMID INFORMATION

The YAC segment Y47G6A was sequenced to span the gap between K03E5
and T28F2. The clone sequenced to the left is K03E5, 200 bp
overlap; the clone sequenced to the right is T28F2, 200 bp overlap.
Actual start of this YAC is at base position 28973 of CELY4E3A;
actual end is at 20751 of CELT28F2.

NOTES:

Coding sequences below are predicted from computer analysis, using
the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES
source

Location/Qualifiers
1..194322
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="I"
/clone="Y47G6A"
complement(706..5939)
/gene="Y47G6A.12"
/note="sep-1"

gene

complement(join(706..1369,1422..1517,1568..1746,
2101..2995,3046..3347,3399..3521,3571..3796,4529..4928,
4986..5145,5196..5939))

CDS

/gene="Y47G6A.12"
/note="contains similarity to cell division-associated
proteins: Emericella nidulans B1MB (GB:M83232) and
Schizosaccharomyces pombe CUT1 (GB:M36179); coded for by
the following C. elegans cDNAs: Yk40c5.3, Yk40c5.5,
Yk47e3.3, Yk127h4.3, Yk127h4.5, Yk134d2.3, Yk134d2.5,
Yk200f10.5, Yk250c9.3, Yk250c9.5, Yk277d4.3, Yk277d4.5,
Yk297e12.5, Yk349c7.5, Yk360b5.3, Yk360b5.5, Yk376d8.3,
Yk376d8.5, Yk416d8.3, Yk416d8.5, Yk429h5.3, Yk429h5.5,
Yk578b6.3, Yk611a4.3, Yk598d10.3, Yk578b6.5, Yk598d10.5"
/codon_start=1
/product="Hypothetical protein Y47G6A.12"
/protein_id="AAF60651.1"
/db_xref="GI:7331963"
/translation="MKITNKSVDKQHIIEKLDLRKNVCTVIGFAETAELQAEISEL
FIAEGVNGPIDMNSLSKLARTSYASSEYFOGLAKYQRTACKRFITWOTLRKAME
CRSKDRFASIPAKLCFFYFNGELCRAYVCLLDYIDLSDDTLAKAALRWMLFGE
TELEIKLKLTWRMDKSKDMFSATFAMNLUKSEYRMELEKMLKLVKSKDPTRS
FSRVELASYVWLCSTLSNPVGSALRECEPDRVSHIQEAAKLSKDSLVNRNIPGLAS
SQFDNSVNASIWPFLDGHQEDSNYYVHIGESTIAMHFEMRECALVNVITAOQDSMSA
MLNLRVALKASFPVLTQNTTAYYSIIIEAGSEKNAKLMRYSCVNLSSNPILIV
RCSTPKGATSRATPNAGSSVSEKONTMRPDLADLLGDLLEDOQSPHPTTRSCVC
NVCYTIPLHSFAAEYMMYSALHSDFSQLSIKHFNDEPARIRERMSQVLMHRDSSV
RPRNITQNEIFGMCVIRWLTKKLSKESADETWEIFNNAUKIVRYLQRTTDMILA
VTQLGDEFMECHNSWMPRTIRKPRVKATIDCAVDILRAVSPGRPRPKVKELKLN
QPFDERFEKVLRLAMNEMNHGILYREWRCRLFYVGRTSRDPEWAAAYAWAESTQI
GARNAQSRLEKCKRGLVTMSGHDRFKTCVQSPDEMTLVQIAMADDTIYLVLKHLAD
RDPITMLAHYSQAVELMDKFTFLDDEMTAKYPGDTIPEEFWRKRIYDGRMTTFV
DEVQKFLGVAASLLMPSGOLGPKAAELAIAIKHLSKGLLGEAKWYQSKLMDAK
SWEALILRFCEMRITDEKFLPLMHRNSVEVMNODDSIVTEKTYTTLVIVCPHLSQF
CWERLPIFEYPIYVGRQVSIHSTFSOLEAMSKQEQIPLQIDVQNAVYILDPDNLGE
TQKRMVYINKFNWEGTVSGAPKSEINEAALSQRDAFFIIGHGSGSVMPRSVLKQST
CNAISLLMCGSVRTIPQALGDFGKTALIDYAMAKPLIVGLVLTWGTGDEIDRLIRM
IDDCFDSKSLTGIDKLQOLSEAMHEARSKARLYLTGAAYVMGLPVYAKQTTPEVE
KDQNLPTPTKTSARTSMRMTVPTPKQEFVTSKSVMPPTPIFSNNENKSPSRAMP
RVLTPTPOVKTFQEEDEAPKKSITTRQLKPLVAPPATPTTTRTTRSSARTPSRNL

gene

6514..15110
/gene="Y47G6A.28"

CDS

join(6514..6585,6645..6750,11554..12015,12506..12599,
12648..13282,14331..14479,14814..15110)
/gene="Y47G6A.28"
/note="coded for by the following C. elegans cDNAs:
CEMSC37F, Yk277a4.5, Yk281g5.5, Yk497f2.5, CEMSC37"
/codon_start=1
/product="Hypothetical protein Y47G6A.28"
/protein_id="AAK95891.1"
/db_xref="GI:15375265"
/translation="MPRIKQAPTKRHSFRMESREADESFLDEIERIEKEQLRVQV
EKKHIOQARLEQSEALTDSNEAARRREPARNLVEGFQVTKERYGNIEHIE
EKYKARESGCPSEVRMDIHNVTYFPTMSAARQOTFVNRFWNTWYKVSAPR
NSISFSDGNRALKSKOLDQOFRRKLRSYDNVDEDDYNGRRGTIISSHVSGIP
LRRSPFSSSKSTAPIKSRHSRSESELDNVFRSPERHSYEDRPTISNAAKEEFS
DEESEATDDESSGDESVLTDEEVEENSNLIRKLKMEIPSPKIQLISLT
LSYTPGLPRNGIKLEFPQIAKLEISTFKPLISSLAEPENKPKNSPYSLKRAY
DSKPKGKSDSPAVSSPITTPVTPRTRLSRAGKPEDNGTEKQNKAPYCKHCELL
KEKKKALEAKNIPSTGTPVVAQTVQKPAKSEKVTPKCOSKRSETVAATPKY
STDSKSKFFLTKKIKLVLFNFCCKNGEKGKVNKAKTRRFSNKKPKVFTKNIQFWT
KIACKVNSLSEFSELSDSDIISIPSSMAVADTELV"

gene

/gene="Y47G6A.13"

CDS

complement(join(6843..6916,7020..7100,7152..7539,
7591..7680,7952..8438,9257..9483,9535..9594))
/gene="Y47G6A.13"
/note="contains similarity to Pfam family PF00069
{eukaryotic protein kinase domain}, score=53.0, E=-1.3e-13,
N=1"

/codon_start=1

/product="Hypothetical protein Y47G6A.13"
/protein_id="AAF60673.1"

/db_xref="GI:7331985"

/translation="MTNSDKSTSSSASKVMRPGHIDIKWRIKALIGKAGCVVYKVED
KNRGYCAAMKVEYDSSEFDRTLQLEVNVLTKLSDRDLKLIDKRRKLYRWYVTTL
CGKLDLMRLMKIQRGDNATAMRVAVFTLYGLKQLHEAGYVHRDYPGNIMTAANKGR
DARVLLIDFGMARTFVMAGDGKKLRLPMRRRIPLRGTIRYCSNNVHERKEQGCDD
LVAMYSIVYLTGLPWSNLKDEKIMSKSTKDLALPELDLDELKSFPEYLRILSY
ADRPNTKYLNLMAATYKLIKFLNFDPEWDEDEDMAKAEKEKEKEKEKEN
KENLSADSMKTAIPAKSAESEVAGEVDEQKVKVDNSSGQQQDTLNVDRKGNKLIS

```

AKDMASSENVNENQYGTAKAGNSVTGDSIDMETATECDQMEYNKKALPRDELQF
VYPAIAPRNFTVEVKIPP"
gene
16733..29478
/genes="Y47G6A.11"
/notes="msh-6"
CDS
join(16733..16868,17929..18209,19439..20203,22406..23211,
24513..24924,25923..26474,27901..28233,29203..29478)
/genes="Y47G6A.11"
/notes="contains similarity to Pfam family PF00488 (DNA
mismatch repair proteins, mutS family), score=296.7,
E=2.9e-85, N=1; coded for by the following C. elegans
cDNAs: yk1b11.3, yk1b11.5, yk8g10.3, yk8g10.5, yk44a10.5,
yk47b4.3, yk47b4.5, yk151b12.3, yk170g6.5, yk170g6.3,
yk290a8.3, yk290a8.5, yk321b5.3, yk321b5.5, yk563a9.5,
yk788e06.5, yk788e06.3, yk563a9.3"
/codon_start=1
/product="Hypothetical protein Y47G6A.11"
/protein_id="AAK95890.1"
/db_xref="GI:15375264"
/translation="MSKROSLMSFFTKPKSEKPEEVKEKSVPEKPKSLKNDTPKIS

Query Match      80.0%; Score 16.8; DB 3; Length 194322;
Best Local Similarity 90.0%; Pred. No. 7.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atgtcggcggtgaaatat 20
|||||
Db 68177 ATGTGGCGCTGGAACTAT 68196

RESULT 12
AC006782/c 319857 bp DNA linear HTG 25-FEB-1999
LOCUS
DEFINITION
Caenorhabditis elegans clone Y47G6, *** SEQUENCING IN PROGRESS ***,
4 unordered pieces.
AC006782.2 GI:4309789
VERSION
AC006782.2
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
Caenorhabditis elegans.
ORGANISM
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.
1 (bases 1 to 319857)
Waterston,R.H.
The sequence of Caenorhabditis elegans clone
Unpublished
2 (bases 1 to 319857)
Waterston,R.H.
Direct Submission
Submitted (23-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 1, 1999 this sequence version replaced gi:4263156.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 3652: contig of 3652 bp in length
* 3653 3666: gap of unknown length
* 3667 36798: contig of 33132 bp in length
* 36799 36812: gap of unknown length
* 36813 156230: contig of 119418 bp in length
* 156231 156244: gap of unknown length
* 156245 319857: contig of 163613 bp in length.
Location/Qualifiers
1. 319857
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y47G6"

BASE COUNT 102084 a 56807 c 58334 g 102590 t 42 others
ORIGIN
Query Match      80.0%; Score 16.8; DB 2; Length 319857;
Best Local Similarity 90.0%; Pred. No. 7.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atgtcggcggtgaaatat 20
|||||
Db 269811 ATGTGGCGCTGGAACTAT 269792

RESULT 13
AC015769
LOCUS
DEFINITION
Homo sapiens clone RP11-2D9, LOW-PASS SEQUENCE SAMPLING.
AC015769
AC015769.3 GI:10998966
VERSION
AC015769.3
KEYWORDS
HTG; HTGS_PHASE0.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 57013)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Unpublished
2 (bases 1 to 57013)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boquslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,B., Gage,D.,
Gallagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 25, 2000 this sequence version replaced gi:9653135.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1509
Center clone name: 2_D_9
* NOTE: This record contains 72 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 662: contig of 662 bp in length
* 663 762: gap of 100 bp
* 763 1465: contig of 703 bp in length

```

* 1466 1565: gap of 100 bp
* 1566 2258: contig of 693 bp in length
* 2259 2358: gap of 100 bp
* 2359 3036: contig of 678 bp in length
* 3037 3136: gap of 100 bp
* 3137 3826: contig of 690 bp in length
* 3827 3926: gap of 100 bp
* 3927 4615: contig of 689 bp in length
* 4616 4715: gap of 100 bp
* 4716 5405: contig of 690 bp in length
* 5406 5505: gap of 100 bp
* 5506 6194: contig of 689 bp in length
* 6195 6294: gap of 100 bp
* 6295 7005: contig of 711 bp in length
* 7006 7105: gap of 100 bp
* 7106 7803: contig of 698 bp in length
* 7804 7903: gap of 100 bp
* 7904 8593: contig of 690 bp in length
* 8594 8693: gap of 100 bp
* 8694 9393: contig of 700 bp in length
* 9394 9493: gap of 100 bp
* 9494 10189: contig of 696 bp in length
* 10190 10289: gap of 100 bp
* 10290 10985: contig of 696 bp in length
* 10986 11085: gap of 100 bp
* 11086 11772: contig of 687 bp in length
* 11773 11872: gap of 100 bp
* 11873 12564: contig of 692 bp in length
* 12565 12664: gap of 100 bp
* 12665 13371: contig of 707 bp in length
* 13372 13471: gap of 100 bp
* 13472 14172: contig of 701 bp in length
* 14173 14272: gap of 100 bp
* 14273 14959: contig of 687 bp in length
* 14960 15059: gap of 100 bp
* 15060 15741: contig of 682 bp in length
* 15742 15841: gap of 100 bp
* 15842 16533: contig of 692 bp in length
* 16534 16633: gap of 100 bp
* 16634 17326: contig of 693 bp in length
* 17327 17426: gap of 100 bp
* 17427 18121: contig of 695 bp in length
* 18122 18221: gap of 100 bp
* 18222 18932: contig of 711 bp in length
* 18933 19032: gap of 100 bp
* 19033 19734: contig of 702 bp in length
* 19735 19834: gap of 100 bp
* 19835 20517: contig of 683 bp in length
* 20518 20617: gap of 100 bp
* 20618 21314: contig of 697 bp in length
* 21315 21414: gap of 100 bp
* 21415 22122: contig of 708 bp in length
* 22123 22222: gap of 100 bp
* 22223 22912: contig of 690 bp in length
* 22913 23012: gap of 100 bp
* 23013 23710: contig of 698 bp in length
* 23711 23810: gap of 100 bp
* 23811 24509: contig of 699 bp in length
* 24510 24609: gap of 100 bp
* 24610 25298: contig of 689 bp in length
* 25299 25398: gap of 100 bp
* 25399 26091: contig of 693 bp in length
* 26092 26191: gap of 100 bp
* 26192 26895: contig of 704 bp in length
* 26896 26995: gap of 100 bp
* 26996 27687: contig of 692 bp in length
* 27688 27787: gap of 100 bp
* 27788 28407: contig of 620 bp in length
* 28408 28507: gap of 100 bp
* 28508 29209: contig of 702 bp in length
* 29210 29309: gap of 100 bp
* 29310 29996: contig of 687 bp in length
* 29997 30096: gap of 100 bp

* 30097 30805: contig of 709 bp in length
* 30806 30905: gap of 100 bp
* 30906 31594: contig of 689 bp in length
* 31595 31694: gap of 100 bp
* 31695 32380: contig of 686 bp in length
* 32381 32480: gap of 100 bp
* 32481 33170: contig of 690 bp in length
* 33171 33270: gap of 100 bp
* 33271 33970: contig of 700 bp in length
* 33971 34070: gap of 100 bp
* 34071 34757: contig of 687 bp in length
* 34758 34857: gap of 100 bp
* 34858 35545: contig of 688 bp in length
* 35546 35645: gap of 100 bp
* 35646 36325: contig of 680 bp in length
* 36326 36425: gap of 100 bp
* 36426 37126: contig of 701 bp in length
* 37127 37226: gap of 100 bp
* 37227 37920: contig of 694 bp in length
* 37921 38020: gap of 100 bp
* 38021 38711: contig of 691 bp in length
* 38712 38811: gap of 100 bp
* 38812 39501: contig of 690 bp in length
* 39502 39601: gap of 100 bp
* 39602 40281: contig of 680 bp in length
* 40282 40381: gap of 100 bp
* 40382 41083: contig of 702 bp in length
* 41084 41183: gap of 100 bp
* 41184 41881: contig of 698 bp in length
* 41882 41981: gap of 100 bp
* 41982 42681: contig of 700 bp in length
* 42682 42781: gap of 100 bp
* 42782 43464: contig of 683 bp in length
* 43465 43564: gap of 100 bp
* 43565 44243: contig of 679 bp in length
* 44244 44343: gap of 100 bp
* 44344 45036: contig of 693 bp in length
* 45037 45136: gap of 100 bp
* 45137 45822: contig of 686 bp in length
* 45823 45922: gap of 100 bp
* 45923 46641: contig of 719 bp in length
* 46642 46741: gap of 100 bp
* 46742 47433: contig of 692 bp in length
* 47434 47533: gap of 100 bp
* 47534 48228: contig of 695 bp in length
* 48229 48328: gap of 100 bp
* 48329 49006: contig of 678 bp in length
* 49007 49106: gap of 100 bp
* 49107 49796: contig of 690 bp in length
* 49797 49896: gap of 100 bp
* 49897 50589: contig of 693 bp in length
* 50590 50689: gap of 100 bp
* 50690 51377: contig of 688 bp in length
* 51378 51477: gap of 100 bp
* 51478 52172: contig of 695 bp in length
* 52173 52272: gap of 100 bp
* 52273 52987: contig of 715 bp in length
* 52988 53087: gap of 100 bp
* 53088 53803: contig of 716 bp in length
* 53804 53903: gap of 100 bp
* 53904 54593: contig of 690 bp in length
* 54594 54693: gap of 100 bp
* 54694 55397: contig of 704 bp in length
* 55398 55497: gap of 100 bp
* 55498 56204: contig of 707 bp in length
* 56205 56304: gap of 100 bp
* 56305 57013: contig of 709 bp in length.

FEATURES Location/Qualifiers
1. .57013

Query Match 78.1%; Score 16.4; DB 2; Length 57013;
Best Local Similarity 94.4%; Pred. No. 1.3e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 tcggcggtggaatac 21
 Db 28531 TCGGGGTGGAATAC 28548

RESULT 14

AC091196
 LOCUS
 DEFINITION Homo sapiens chromosome 11 clone RP11-371C18 map 11, LOW-PASS
 SEQUENCE SAMPLING.

AC091196
 VERSION
 KEYWORDS
 SOURCE HTGS_PHASE0.
 human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

1 (bases 1 to 63528)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A.,
 Camarata, J., Campiano, A., Chang, J., Choepel, Y., Colangelo, M.,
 Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K.,
 Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W.,
 Iliev, I., Johnson, R., Jones, C., Karatas, A., Levine, R., Liu, G.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M.,
 McEwan, P., McKernan, K., McPheeters, B., Meldrum, J., Meneus, L.,
 Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,
 Severy, P., Sougnez, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S.,
 Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (03-APR-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L13097
 Center clone name: 371_C_18

* NOTE: This record contains 77 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 * 1 723: contig of 723 bp in length
 * 724 823: gap of 100 bp
 * 824 1583: contig of 760 bp in length

1584 1683: gap of 100 bp
 1684 2419: contig of 736 bp in length
 2420 2519: gap of 100 bp
 2520 3229: contig of 710 bp in length
 3230 3329: gap of 100 bp
 3330 4081: contig of 752 bp in length
 4082 4181: gap of 100 bp
 4182 4906: contig of 725 bp in length
 4907 5006: gap of 100 bp
 5007 5728: contig of 722 bp in length
 5729 5828: gap of 100 bp
 5829 6537: contig of 709 bp in length
 6538 6637: gap of 100 bp
 6638 7399: contig of 762 bp in length
 7400 7499: gap of 100 bp
 7500 8234: contig of 735 bp in length
 8235 8334: gap of 100 bp
 8335 9048: contig of 714 bp in length
 9049 9148: gap of 100 bp
 9149 9860: contig of 712 bp in length
 9861 9960: gap of 100 bp
 9961 10675: contig of 715 bp in length
 10676 10775: gap of 100 bp
 10776 11495: contig of 720 bp in length
 11496 11595: gap of 100 bp
 11596 12316: contig of 721 bp in length
 12317 12416: gap of 100 bp
 12417 13125: contig of 709 bp in length
 13126 13225: gap of 100 bp
 13226 13949: contig of 724 bp in length
 13950 14049: gap of 100 bp
 14050 14788: contig of 739 bp in length
 14789 14888: gap of 100 bp
 14889 15563: contig of 675 bp in length
 15564 15663: gap of 100 bp
 15664 16408: contig of 745 bp in length
 16409 16508: gap of 100 bp
 16509 17277: contig of 769 bp in length
 17278 17377: gap of 100 bp
 17378 18070: contig of 693 bp in length
 18071 18170: gap of 100 bp
 18171 18875: contig of 705 bp in length
 18876 18975: gap of 100 bp
 18976 19686: contig of 711 bp in length
 19687 19786: gap of 100 bp
 19787 20530: contig of 744 bp in length
 20531 20630: gap of 100 bp
 20631 21361: contig of 731 bp in length
 21362 21461: gap of 100 bp
 21462 22196: contig of 735 bp in length
 22197 22296: gap of 100 bp
 22297 23007: contig of 711 bp in length
 23008 23107: gap of 100 bp
 23108 23842: contig of 735 bp in length
 23843 23942: gap of 100 bp
 23943 24717: contig of 775 bp in length
 24718 24817: gap of 100 bp
 24818 25588: contig of 771 bp in length
 25589 25688: gap of 100 bp
 25689 26399: contig of 711 bp in length
 26400 26499: gap of 100 bp
 26500 27210: contig of 711 bp in length
 27211 27310: gap of 100 bp
 27311 28044: contig of 734 bp in length
 28045 28144: gap of 100 bp
 28145 28859: contig of 715 bp in length
 28860 28959: gap of 100 bp
 28960 29671: contig of 712 bp in length
 29672 29771: gap of 100 bp
 29772 30508: contig of 737 bp in length
 30509 30608: gap of 100 bp
 30609 31354: contig of 746 bp in length
 31355 31454: gap of 100 bp

* 31455 32147: contig of 693 bp in length
* 32148 32247: gap of 100 bp
* 32248 32975: contig of 728 bp in length
* 32976 33075: gap of 100 bp
* 33076 33804: contig of 729 bp in length
* 33805 33904: gap of 100 bp
* 33905 34665: contig of 761 bp in length
* 34666 34765: gap of 100 bp
* 34766 35467: contig of 702 bp in length
* 35468 35567: gap of 100 bp
* 35568 36316: contig of 749 bp in length
* 36317 36416: gap of 100 bp
* 36417 37129: contig of 713 bp in length
* 37130 37229: gap of 100 bp
* 37230 37953: contig of 724 bp in length
* 37954 38053: gap of 100 bp
* 38054 38763: contig of 710 bp in length
* 38764 38863: gap of 100 bp
* 38864 39582: contig of 719 bp in length
* 39583 39682: gap of 100 bp
* 39683 40424: contig of 742 bp in length
* 40425 40524: gap of 100 bp
* 40525 41236: contig of 712 bp in length
* 41237 41336: gap of 100 bp
* 41337 42037: contig of 701 bp in length
* 42038 42137: gap of 100 bp
* 42138 42869: contig of 732 bp in length
* 42870 42969: gap of 100 bp
* 42970 43660: contig of 691 bp in length
* 43661 43760: gap of 100 bp
* 43761 44460: contig of 700 bp in length
* 44461 44560: gap of 100 bp
* 44561 45293: contig of 733 bp in length
* 45294 45393: gap of 100 bp
* 45394 46142: contig of 749 bp in length
* 46143 46242: gap of 100 bp
* 46243 46941: contig of 699 bp in length
* 46942 47041: gap of 100 bp
* 47042 47821: contig of 780 bp in length
* 47822 47921: gap of 100 bp
* 47922 48704: contig of 783 bp in length
* 48705 48804: gap of 100 bp
* 48805 49564: contig of 760 bp in length
* 49565 49664: gap of 100 bp
* 49665 50364: contig of 700 bp in length
* 50365 50464: gap of 100 bp
* 50465 51179: contig of 715 bp in length
* 51180 51279: gap of 100 bp
* 51280 52001: contig of 722 bp in length
* 52002 52101: gap of 100 bp
* 52102 52822: contig of 721 bp in length
* 52823 52922: gap of 100 bp
* 52923 53650: contig of 728 bp in length
* 53651 53750: gap of 100 bp
* 53751 54459: contig of 709 bp in length
* 54460 54559: gap of 100 bp
* 54560 55296: contig of 737 bp in length
* 55297 55396: gap of 100 bp
* 55397 56105: contig of 709 bp in length
* 56106 56205: gap of 100 bp
* 56206 56968: contig of 763 bp in length
* 56969 57068: gap of 100 bp

Query Match 78.1%; Score 16.4; DB 2; Length 63528;
Best Local Similarity 94.4%; Pred. No. 1.3e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 tcggcggtggaatc 21
||||| |||||||
DB 34789 TCGGGGTGGAATC 34806

RESULT 15

AC010774
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT

AC010774
Homo sapiens clone RP11-2G12, LOW-PASS SEQUENCE SAMPLING.
AC010774
HTG: HTGS_PHASE0.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 100511)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-2G12
Unpublished
2 (bases 1 to 100511)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:5910818.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2531
Center clone name: 2_G_12

* NOTE: This record contains 105 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 846: contig of 846 bp in length
* 847 946: gap of 100 bp
* 947 1762: contig of 816 bp in length
* 1763 1862: gap of 100 bp
* 1863 2715: contig of 853 bp in length
* 2716 2815: gap of 100 bp
* 2816 3678: contig of 863 bp in length
* 3679 3778: gap of 100 bp
* 3779 4598: contig of 820 bp in length
* 4599 4698: gap of 100 bp
* 4699 5538: contig of 840 bp in length
* 5539 5638: gap of 100 bp
* 5639 6483: contig of 845 bp in length
* 6484 6583: gap of 100 bp
* 6584 7400: contig of 817 bp in length
* 7401 7500: gap of 100 bp in length
* 7501 8349: contig of 849 bp in length

Thu Jul 25 08:38:03 2002

* 8350 8449: gap of 100 bp
* 8450 9306: contig of 857 bp in length
* 9307 9406: gap of 100 bp
* 9407 10280: contig of 874 bp in length
* 10281 10380: gap of 100 bp
* 10381 11235: contig of 855 bp in length
* 11236 11335: gap of 100 bp
* 11336 12194: contig of 859 bp in length
* 12195 12294: gap of 100 bp
* 12295 13133: contig of 839 bp in length
* 13134 13233: gap of 100 bp
* 13234 14096: contig of 863 bp in length
* 14097 14196: gap of 100 bp
* 14197 15064: contig of 868 bp in length
* 15065 15164: gap of 100 bp
* 15165 16038: contig of 874 bp in length
* 16039 16138: gap of 100 bp
* 16139 16989: contig of 851 bp in length
* 16990 17089: gap of 100 bp
* 17090 17928: contig of 839 bp in length
* 17929 18028: gap of 100 bp
* 18029 18849: contig of 821 bp in length
* 18850 18949: gap of 100 bp
* 18950 19797: contig of 848 bp in length
* 19798 19897: gap of 100 bp
* 19898 20763: contig of 866 bp in length
* 20764 20863: gap of 100 bp
* 20864 21761: contig of 898 bp in length
* 21762 21861: gap of 100 bp
* 21862 22725: contig of 864 bp in length
* 22726 22825: gap of 100 bp
* 22826 23705: contig of 880 bp in length
* 23706 23805: gap of 100 bp
* 23806 24664: contig of 859 bp in length
* 24665 24764: gap of 100 bp
* 24765 25614: contig of 850 bp in length
* 25615 25714: gap of 100 bp
* 25715 26555: contig of 841 bp in length
* 26556 26655: gap of 100 bp
* 26656 27508: contig of 853 bp in length
* 27509 27608: gap of 100 bp
* 27609 28472: contig of 864 bp in length
* 28473 28572: gap of 100 bp
* 28573 29439: contig of 867 bp in length
* 29440 29539: gap of 100 bp
* 29540 30406: contig of 867 bp in length
* 30407 30506: gap of 100 bp
* 30507 31382: contig of 876 bp in length
* 31383 31482: gap of 100 bp
* 31483 32342: contig of 860 bp in length
* 32343 32442: gap of 100 bp
* 32443 33319: contig of 877 bp in length
* 33320 33419: gap of 100 bp
* 33420 34276: contig of 857 bp in length
* 34277 34376: gap of 100 bp
* 34377 35258: contig of 882 bp in length
* 35259 35358: gap of 100 bp
* 35359 36218: contig of 860 bp in length
* 36219 36318: gap of 100 bp
* 36319 37161: contig of 843 bp in length
* 37162 37261: gap of 100 bp
* 37262 38153: contig of 892 bp in length
* 38154 38253: gap of 100 bp
* 38254 39076: contig of 823 bp in length
* 39077 39176: gap of 100 bp
* 39177 40044: contig of 868 bp in length
* 40045 40144: gap of 100 bp
* 40145 41010: contig of 866 bp in length
* 41011 41110: gap of 100 bp
* 41111 42014: contig of 904 bp in length
* 42015 42114: gap of 100 bp
* 42115 42982: contig of 868 bp in length
* 42983 43082: gap of 100 bp

* 43083 43950: contig of 868 bp in length
* 43951 44050: gap of 100 bp
* 44051 44917: contig of 867 bp in length
* 44918 45017: gap of 100 bp
* 45018 45895: contig of 878 bp in length
* 45896 45995: gap of 100 bp
* 45996 46867: contig of 872 bp in length
* 46868 46967: gap of 100 bp
* 46968 47866: contig of 899 bp in length
* 47867 47966: gap of 100 bp
* 47967 48832: contig of 866 bp in length
* 48833 48932: gap of 100 bp
* 48933 49799: contig of 867 bp in length
* 49800 49899: gap of 100 bp
* 49900 50753: contig of 854 bp in length
* 50754 50853: gap of 100 bp
* 50854 51718: contig of 865 bp in length
* 51719 51818: gap of 100 bp
* 51819 52705: contig of 887 bp in length
* 52706 52805: gap of 100 bp
* 52806 53655: contig of 850 bp in length
* 53656 53755: gap of 100 bp
* 53756 54592: contig of 837 bp in length
* 54593 54692: gap of 100 bp
* 54693 55567: contig of 875 bp in length
* 55568 55667: gap of 100 bp
* 55668 56545: contig of 878 bp in length
* 56546 56645: gap of 100 bp
* 56646 57481: contig of 836 bp in length
* 57482 57581: gap of 100 bp
* 57582 58427: contig of 846 bp in length
* 58428 58527: gap of 100 bp
* 58528 59378: contig of 851 bp in length
* 59379 59478: gap of 100 bp
* 59479 60348: contig of 870 bp in length
* 60349 60448: gap of 100 bp
* 60449 61307: contig of 859 bp in length
* 61308 61407: gap of 100 bp
* 61408 62279: contig of 872 bp in length
* 62280 62379: gap of 100 bp
* 62380 63208: contig of 829 bp in length
* 63209 63308: gap of 100 bp
* 63309 64163: contig of 855 bp in length
* 64164 64263: gap of 100 bp
* 64264 65109: contig of 846 bp in length
* 65110 65209: gap of 100 bp
* 65210 66079: contig of 870 bp in length
* 66080 66179: gap of 100 bp
* 66180 67031: contig of 852 bp in length
* 67032 67131: gap of 100 bp
* 67132 67974: contig of 843 bp in length
* 67975 68074: gap of 100 bp
* 68075 68904: contig of 830 bp in length
* 68905 69004: gap of 100 bp
* 69005 69857: contig of 853 bp in length

Query Match 78.1%; Score 16.4; DB 2; Length 100511;
Best Local Similarity 94.4%; Pred. No. 1.3e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 tgcggcggtggaaata 19
| | | | | | | | | | | | | | | | | |
Db 47835 TGTGCGGCGGTGGAAAAA 47852

Search completed: July 25, 2002, 05:38:07
Job Time: 6640 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 05:44:40 ; Search time 378.07 Seconds
(without alignments)
95.366 Million cell updates/sec

Title: US-09-235-416-3

Perfect score: 21
Sequence: 1 atgtcggcggtggaatc 21

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21	100.0	21	20	AXX87657	Kinesin motor prot
2	21	100.0	20	20	AXX87656	Thermomyces lanugi
C 3	16.2	77.1	2707	22	AAS44925	cDNA encoding nove
C 4	15.8	75.2	152	21	AAA45638	Human secreted exp
5	15.8	75.2	704	21	AAA01667	Human colon cancer
6	15.8	75.2	1209	21	AAA64407	Open reading frame
7	15.8	75.2	1449	21	AAA62003	Hydrophobic domain
8	15.8	75.2	1652	21	AAA62013	Hydrophobic domain
9	15.8	75.2	1793	22	AAH14868	Human cDNA sequenc

10	15.8	75.2	2252	21	AAA64406	DNA encoding a hum	
11	15.8	75.2	2285	22	AAS21290	Human cDNA sequenc	
12	15.8	75.2	2437	22	AAF83002	Human MBSP10 poly	
13	15.8	75.2	2449	22	AAF59608	Human cell cycle a	
14	15.8	75.2	2495	22	AAS45041	cDNA encoding nove	
15	15.8	75.2	2505	21	AAC76426	Human ORFX ORF1981	
16	15.8	75.2	2505	21	AAA23447	cDNA encoding huma	
17	15.8	75.2	2607	22	AAS45229	cDNA encoding nove	
18	15.8	75.2	4623	22	ABA09048	Human ASB-3 protei	
19	15.8	75.2	12393	24	ABL33263	Human immune syste	
20	15.4	73.3	8035	21	AAA51305	Human ASB-3 protei	
21	15.4	73.3	31328	21	AAA51295	A. terreus LovF/Sc	
c	22	15.2	60	20	AAZ19638	A. terreus DAB/lov	
c	23	15.2	72.4	61	AAZ19678	Complement system	
c	24	15.2	72.4	442	21	AAC00237	Complement system
c	25	15.2	72.4	503	21	AAI77917	Human secreted pro
c	26	15.2	72.4	503	22	AAI28655	cDNA encoding huma
c	27	15.2	72.4	503	22	AAI29078	Colon tumour relat
c	28	15.2	72.4	780	23	AAS56003	Colon tumour relat
c	29	15.2	72.4	1319	21	AAI77873	Salmonella typhi D
c	30	15.2	72.4	1425	23	AAI76017	Human cancer assoc
c	31	15.2	72.4	1933	22	AAK85566	DNA encoding novel
c	32	15.2	72.4	2700	22	AAH74536	Human immune/haema
c	33	15.2	72.4	2739	22	AAH74534	A human alpha-2 ma
c	34	15.2	72.4	3227	23	ABL20897	A human alpha-2 ma
c	35	15.2	72.4	3239	23	ABL02077	Drosophila melanog
c	36	15.2	72.4	3246	23	ABL02074	Drosophila melanog
c	37	15.2	72.4	3405	22	AAH74533	A human alpha-2 ma
c	38	15.2	72.4	3413	22	ABA14721	A human alpha-2 ma
c	39	15.2	72.4	4065	22	AAH74535	Human nervous syst
c	40	15.2	72.4	4527	22	AAH74529	DNA encoding a hum
c	41	15.2	72.4	4527	22	AAH74530	A human alpha-2 ma
c	42	15.2	72.4	4734	23	ABL13666	A human alpha-2 ma
c	43	15.2	72.4	4771	22	AAH74532	Drosophila melanog
c	44	15.2	72.4	4771	23	AAS85411	A human alpha-2 ma
c	45	15.2	72.4	5343	23	ABL20896	DNA encoding novel

ALIGNMENTS

RESULT 1
AXX87657
ID AAX87657 standard; DNA; 21 BP.
XX
AC AAX87657;
XX
DT 26-OCT-1999 (first entry)
XX
DE Kinesin motor protein TL-gamma gene PCR primer.
XX
KW TL-gamma; kinesin; motor protein; microtubule; unc-104; infection;
KW neurodegenerative disease; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease;
KW anyotrophic lateral sclerosis; PCR; primer; ss.
XX
OS Synthetic.
OS Thermomyces lanuginosus.
XX
XX WO9937659-A1.
XX
PD 29-JUL-1999.
XX
PF 22-JAN-1999; 99WO-US01355.
XX
PR 23-JAN-1998; 98US-0072361.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goldstein LSB, Sakowicz R;
XX
DR WPI; 1999-493950/41.
XX

PT New nucleic acid encoding microtubule motor protein, used for
PT diagnosis of fungal infection and neurodegenerative disease
XX
XX
PS Claim 7; Page 60; 75pp; English.
XX
CC This oligonucleotide represents one of a claimed primer pair (see
CC also AAX87658) that can be used in the amplification of microtubule
CC motor protein nucleic acids, especially TL-gamma DNA or RNA from a
CC hyphal fungus such as Thermomyces lanuginosus (see AAX87656).
CC TL-gamma (see AAX06618) is a novel ATP-dependent, plus end-directed
CC microtubule motor protein that is a member of the unc-104 family
CC and kinesin superfamily. The invention provides TL-gamma nucleic
CC acids, proteins and antibodies, and methods of screening for
CC TL-gamma modulators potentially useful for treating hyphal fungal
CC infections and diseases caused by mutated TL-gamma, e.g.
CC neurodegeneration. TL-gamma nucleic acids are also used to identify
CC polymorphic variants, alleles, homologues etc. of TL-gamma, or
CC other motor proteins, to generate protein structural models, for
CC recombinant production of TL-gamma, as antisense molecules, to
CC produce transgenic or knockout animals, and in gene therapy.
CC Detection of TL-gamma allows differentiation between hyphal and
CC non-hyphal fungal infections.
XX
SQ Sequence 21 BP; 5 A; 3 C; 8 G; 5 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.24; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0;

QY 1 atgtcggcggtggaataatc 21
|||||
Db 1 atgtcggcggtggaataatc 21

RESULT 2
AAX87656
ID AAX87656 standard; DNA; 2352 BP.
XX
AC AAX87656;
XX
DT 26-OCT-1999 (first entry)
XX
DE Thermomyces lanuginosus kinesin motor protein TL-gamma DNA.
XX
XX TL-gamma; kinesin; motor protein; microtubule; unc-104; infection;
KW neurodegenerative disease; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease;
KW anyotrophic lateral sclerosis; ss.
XX
XX Thermomyces lanuginosus.
XX
XX WO9937659-A1.
XX
XX 29-JUL-1999.
XX
XX 22-JAN-1999; 99WO-US01355.
XX
XX 23-JAN-1998; 98US-0072361.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Goldstein LSB, Sakowicz R;
XX
XX WPI: 1999-493950/41.
XX
XX P-PSDB; AAY06618.
XX
XX New nucleic acid encoding microtubule motor protein, used for
PT diagnosis of fungal infection and neurodegenerative disease
XX
XX Claim 4; Page 69-70; 75pp; English.
XX
XX This is the DNA sequence of the Thermomyces lanuginosus gene

CC coding for TL-gamma (see AAY06618), a novel ATP-dependent, plus
CC end-directed microtubule motor protein that is a member of the
CC unc-104 family and kinesin superfamily. The gene was isolated
CC from genomic DNA by PCR amplification (see also AAX87659-61).
CC The invention provides TL-gamma nucleic acids, proteins and
CC antibodies, and methods of screening for TL-gamma modulators
CC potentially useful for treating hyphal fungal infections and
CC diseases caused by mutated TL-gamma, e.g. neurodegeneration
CC involving anterograde axonal transport, such as Alzheimer's,
CC Parkinson's or Huntington's diseases or amyotrophic lateral
CC sclerosis. TL-gamma nucleic acids are also used to identify
CC polymorphic variants, alleles, homologues etc. of TL-gamma, or
CC other motor proteins, to generate protein structural models, for
CC recombinant production of TL-gamma, as antisense molecules, to
CC produce transgenic or knockout animals, and in gene therapy.
CC Detection of TL-gamma allows differentiation between hyphal and
CC non-hyphal fungal infections.
XX
SQ Sequence 2352 BP; 626 A; 564 C; 672 G; 490 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 2352;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaataatc 21
|||||
Db 1 atgtcggcggtggaataatc 21

RESULT 3
AAS44925/c
ID AAS44925 standard; cDNA; 2707 BP.
XX
XX AAS44925;
XX
XX 18-DEC-2001 (first entry)
XX
DE cDNA encoding novel human secretory protein, Seq ID No 6.
XX
XX Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
KW anyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;
KW gut protection; lung; liver fibrosis; immune deficiency; infection;
KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;
KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
KW fertility; analgesic; pain; antigen; ss.
XX
XX Homo sapiens.
XX
XX WO200166689-A2.
XX
XX 13-SEP-2001.
XX
XX 05-MAR-2001; 2001WO-US04942.
XX
XX 07-MAR-2000; 2000US-0519705.
XX
XX 19-MAY-2000; 2000US-0574454.
XX
XX 17-JUN-2000; 2000US-0596193.
XX
XX 14-JUL-2000; 2000US-0616847.
XX
XX 19-SEP-2000; 2000US-0665363.
XX
XX 20-OCT-2000; 2000US-0693267.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;
PI Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AU, Wang J;
XX
XX WPI: 2001-589934/66.
XX
XX P-PSDB; AAU28025.
XX

KW Human; colon cancer; tumour; diagnosis; gene expression product;
 KW probe; detection; cancerous state; metastasis; identification;
 KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN W09958675-A2.
 XX
 XX 18-NOV-1999.
 XX
 XX 13-MAY-1999; 99WO-US10602.
 XX
 PR 14-MAY-1998; 98US-0085426.
 PR 15-MAY-1998; 98US-0085537.
 PR 15-MAY-1998; 98US-0085596.
 PR 21-OCT-1998; 98US-0105234.
 PR 27-OCT-1998; 98US-0105877.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Lanson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
 XX
 XX WPI: 2000-126369/11.
 DR
 XX Polynucleotide library used to determine cancerous states of mammalian
 PT cells -
 PT
 XX
 PS Claim 1; Page 599-600; 1097pp; English.
 XX
 CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
 CC libraries constructed from human colon cancer cell lines. The present
 CC invention also describes a method of detecting differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell, comprising
 CC detecting at least one differentially expressed gene product in a test
 CC sample derived from a cell suspected of being cancerous, where detection
 CC of the differentially expressed gene product is correlated with a
 CC cancerous state of the cell from which the test sample was derived.
 CC The polynucleotide sequences can be used in a method for detecting
 CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
 CC negative breast cancer, lung cancer, and colon cancer.
 XX
 SQ Sequence 704 BP; 217 A; 134 C; 153 G; 194 T; 6 other;
 Query Match 75.2%; Score 15.8; DB 21; Length 704;
 Best Local Similarity 89.5%; Pred. No. 1.4e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 tgcggcggtggaatat 20
 ||||| |||||
 Db 214 tgcgggttggtggaatat 232
 RESULT 6
 AAA64407
 ID AAA64407 standard; cDNA; 1209 BP.
 XX
 AC AAA64407;
 XX
 XX 02-JAN-2001 (first entry)
 DT
 XX Open reading frame of DNA encoding a human TANGO 201 polypeptide.

XX TANGO 201; TANGO 223; secreted protein; transmembrane protein;
 KW proliferative disorder; neoplasm; tumour; pancreatic disorder; cirrhosis;
 KW pancreatitis; thyroiditis; goiter; Graves' disease; gastric disorder;
 KW gastritis; tumour; placental disorder; placentalitis; spontaneous abortion;
 KW pulmonary disorder; atelectasis; edema; Goodpasture's syndrome;
 KW muscular dystrophy; cardiovascular disorder; ischemic heart disease;
 KW congenital heart disease; cerebral edema; cerebrovascular disease;
 KW hepatic disorder; jaundice; hepatitis; leukocytic disorder; hyperplasia;
 KW leukopenia; leukocytosis; malignant lymphoma; prostate disorder;
 KW inflammatory disease; ss.
 XX
 OS Homo sapiens.
 XX
 XX Location/Qualifiers
 FH 1..1209
 FT /*tag= a "TANGO 201"
 FT /product= "TANGO 201"
 FT sig_peptide 1..99
 FT /*tag= b
 FT mat_peptide 100..1209
 FT /*tag= c
 FT
 XX W0200050442-A2.
 PN 31-AUG-2000.
 XX
 XX 25-FEB-2000; 2000WO-US04784.
 XX
 XX 26-FEB-1999; 99US-0259388.
 PR
 XX (MILL-) MILLENNIUM PHARM INC.
 PA
 PI McCarthy SA;
 PI
 XX WPI: 2000-572066/53.
 DR P-PSDB; AAB08640.
 XX
 PT Novel nucleic acid molecule encoding secreted or transmembrane protein
 PT useful for identifying modulators and for diagnosing and treating
 PT pancreatic, cardiovascular, liver and pituitary disorders -
 XX
 PS Claim 2; Page 120-121; 176pp; English.
 XX
 CC The present sequence encodes a TANGO 201 polypeptide. The specification
 CC also describes a TANGO 223 polypeptide. These polypeptides are secreted
 CC or transmembrane proteins. Human TANGO 201 and 223 nucleic acids,
 CC proteins and their modulators are useful for treating proliferative
 CC disorders e.g. neoplasms or tumours, pancreatic disorders
 CC (e.g. pancreatitis), disorders of the adrenal cortex, adrenal medulla,
 CC thyroid gland (e.g. thyroiditis), goiter, Graves' disease, gastric
 CC disorders (e.g. gastritis or tumours), placental disorders (e.g.
 CC placentalitis or spontaneous abortion), pulmonary disorders (e.g.
 CC atelectasis), edema, Goodpasture's syndrome, disorders of the skeletal
 CC muscle (e.g. muscular dystrophy), cardiovascular disorders (e.g.
 CC ischemic heart disease and congenital heart disease), disorders of the
 CC brain (e.g. cerebral edema), cerebrovascular disease and to treat
 CC injury or trauma to the brain. They are also useful for treating
 CC hepatic disorders (e.g. jaundice, hepatitis, cirrhosis or malignant
 CC tumours), renal, testicular, intestinal disorders. TANGO 223
 CC polynucleotides are also useful for treating leukocytic disorders (e.g.
 CC leukopenias, leukocytosis and malignant lymphomas) and prostate
 CC disorders (e.g. inflammatory diseases, hyperplasia or tumours).
 XX
 SQ Sequence 1209 BP; 380 A; 247 C; 282 G; 300 T; 0 other;

Query Match 75.2%; Score 15.8; DB 21; Length 1209;
 Best Local Similarity 89.5%; Pred. No. 1.4e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 tgcggcggtggaatat 20
 ||||| |||||
 Db 214 tgcgggttggtggaatat 232

Db	882	tgtcgggttggtggaatat	900
RESULT	7		
AAA62003			
ID	AAA62003	standard; DNA; 1449 BP.	
XX	AAAG2003;		
AC	AAAG2003;		
DT	02-FEB-2001	(first entry)	
XX			
DE	Hydrophobic domain protein cDNA HP03091 isolated from Liver cells.		
XX			
KW	Human; secreted protein; membrane protein; hydrophobic domain; proliferation control; differentiation induction; material transport; biophylaxis; signal receptor; ion channel; transporter; immunostimulant; immunosuppressant; haematopoiesis regulator; chemotactic; chemokinetic; haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; autoimmune disease; Alzheimer's disease; Parkinson's disease; cancer; gene therapy; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200029448-A2.		
XX			
PD	25-MAY-2000.		
XX			
PF	17-NOV-1999;	99WO-JP06412.	
XX			
PR	17-NOV-1998;	98JP-0326255.	
PR	22-DEC-1998;	98JP-0364315.	
PR	16-MAR-1999;	99JP-0069811.	
PR	27-APR-1999;	99JP-0119299.	
PR	19-MAY-1999;	99JP-0138169.	
XX			
PA	(SAGA) SAGAMI CHEM RES CENT.		
PA	(PROT-) PROTEGENE INC.		
XX			
PI	Kato S, Kimura T;		
XX			
DR	WPI: 2000-387753/33.		
DR	P-PSDB; AAB12130.		
XX			
PT	Proteins comprising hydrophobic regions, such as secretory and membrane proteins, useful in research and diagnostics and having various activities e.g. immunomodulatory, antiinflammatory, chemokinetic, hemostatic, thrombolytic -		
XX			
PS	Claim 3; Page 248-249; 410pp; English.		
XX			
CC	Secretory proteins play important roles in the proliferation control, the differentiation induction, the material transport and the biophylaxis of cells. Membrane proteins have important roles as signal receptors, ion channels and transporters. The present sequence is the coding sequence for a human protein which has at least one hydrophobic domain. The protein encoded by the present sequence may be a secretory or a membrane protein. The encoded protein may have cytokine and cell proliferation/differentiation activity, immune stimulating or suppressing activity, haematopoiesis activity, tissue growth activity, haemostatic activity/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, anti-inflammatory activity and tumour inhibition activity. The present sequence could therefore be used for treatment of autoimmune disease, Alzheimer's disease, Parkinson's disease, and cancer via gene therapy.		
XX			
SQ	Sequence 1449 BP; 458 A; 292 C; 333 G; 366 T; 0 other;		
Query Match	75.2%;	Score 15.8; DB 21; Length 1449;	
Best Local Similarity	89.5%;	Pred. NO. 1.4e+02;	
Matches	17; Conservative	0; Mismatches 2; Indels 0; Gaps 0;	
QY	2	tgtcggcggtggaatat	20

Db	1044	tgtcgggttggtggaatat	1062
RESULT	8		
AAA62013			
ID	AAA62013	standard; DNA; 1652 BP.	
XX			
AC	AAAG2013;		
XX			
DT	02-FEB-2001	(first entry)	
XX			
DE	Hydrophobic domain protein cDNA HP03091 isolated from Liver cells.		
XX			
KW	Human; secreted protein; membrane protein; hydrophobic domain; proliferation control; differentiation induction; material transport; biophylaxis; signal receptor; ion channel; transporter; immunostimulant; immunosuppressant; haematopoiesis regulator; chemotactic; chemokinetic; haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; autoimmune disease; Alzheimer's disease; Parkinson's disease; cancer; gene therapy; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200029448-A2.		
XX			
PD	25-MAY-2000.		
XX			
PF	17-NOV-1999;	99WO-JP06412.	
XX			
PR	17-NOV-1998;	98JP-0326255.	
PR	22-DEC-1998;	98JP-0364315.	
PR	16-MAR-1999;	99JP-0069811.	
PR	27-APR-1999;	99JP-0119299.	
PR	19-MAY-1999;	99JP-0138169.	
XX			
PA	(SAGA) SAGAMI CHEM RES CENT.		
PA	(PROT-) PROTEGENE INC.		
XX			
PI	Kato S, Kimura T;		
XX			
DR	WPI: 2000-387753/33.		
DR	P-PSDB; AAB12130.		
XX			
PT	Proteins comprising hydrophobic regions, such as secretory and membrane proteins, useful in research and diagnostics and having various activities e.g. immunomodulatory, antiinflammatory, chemokinetic, hemostatic, thrombolytic -		
XX			
PS	Claim 4; Page 267-270; 410pp; English.		
XX			
CC	Secretory proteins play important roles in the proliferation control, the differentiation induction, the material transport and the biophylaxis of cells. Membrane proteins have important roles as signal receptors, ion channels and transporters. The present sequence is the coding sequence for a human protein which has at least one hydrophobic domain. The protein encoded by the present sequence may be a secretory or a membrane protein. The encoded protein may have cytokine and cell proliferation/differentiation activity, immune stimulating or suppressing activity, haematopoiesis activity, tissue growth activity, haemostatic activity/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, anti-inflammatory activity and tumour inhibition activity. The present sequence could therefore be used for treatment of autoimmune disease, Alzheimer's disease, Parkinson's disease, and cancer via gene therapy.		
XX			
SQ	Sequence 1652 BP; 526 A; 323 C; 379 G; 424 T; 0 other;		
Query Match	75.2%;	Score 15.8; DB 21; Length 1652;	
Best Local Similarity	89.5%;	Pred. NO. 1.5e+02;	
Matches	17; Conservative	0; Mismatches 2; Indels 0; Gaps 0;	

CC injury or trauma to the brain. They are also useful for treating
 CC hepatic disorders (e.g. jaundice, hepatitis, cirrhosis or malignant
 CC tumours), renal, testicular, intestinal disorders. TANGO 223
 CC polynucleotides are also useful for treating leukocytic disorders (e.g.
 CC leukopenias, leukocytosis and malignant lymphomas) and prostate
 CC disorders (e.g. inflammatory diseases, hyperplasia or tumours).
 XX Sequence 2252 BP; 681 A; 444 C; 496 G; 631 T; 0 other;

Query Match 75.2%; Score 15.8; DB 21; Length 2252;
 Best Local Similarity 89.5%; Pred. No. 1.5e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgctggcggtggaatat 20
 ||||| |||||
 Db 1060 tgctggttggtggaatat 1078

RESULT 11
 AAS21290
 ID AAS21290 standard; cDNA; 2285 BP.
 AC AAS21290;
 XX AAS21290;

DT 24-OCT-2001 (first entry)

DE Human cDNA sequence encoding for PRO4321 polypeptide.

XX Human secretory and transmembrane; PRO; mammalian; cancer; lung;
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
 KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.
 XX Homo sapiens.

OS Homo sapiens.

XX WO200140466-A2.

PN WO200140466-A2.

XX 07-JUN-2001.

PD 01-DEC-2000; 2000WO-US32678.

XX 01-DEC-1999; 99WO-US28301.

PR 01-DEC-1999; 99WO-US28634.

PR 02-DEC-1999; 99WO-US28551.

PR 02-DEC-1999; 99WO-US28564.

PR 02-DEC-1999; 99WO-US28565.

PR 09-DEC-1999; 99US-0170262.

PR 16-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30911.

PR 20-DEC-1999; 99WO-US30999.

PR 30-DEC-1999; 99WO-US31243.

PR 06-JAN-2000; 2000WO-US00277.

PR 06-JAN-2000; 2000WO-US03376.

PR 11-FEB-2000; 2000WO-US03565.

PR 18-FEB-2000; 2000WO-US04341.

PR 18-FEB-2000; 2000WO-US04342.

PR 22-FEB-2000; 2000WO-US04414.

PR 24-FEB-2000; 2000WO-US04914.

PR 24-FEB-2000; 2000WO-US05004.

PR 01-MAR-2000; 2000WO-US05601.

PR 21-MAR-2000; 2000WO-US07377.

PR 21-MAR-2000; 2000WO-US07532.

PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 XX WPI; 2001-403281/43.
 DR P-PSDB; AAU12218.
 XX Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical -
 XX Claim 3; Fig 93; 813pp; English.

PS AAS21244-AAS21518 encode for novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.
 XX Sequence 2285 BP; 693 A; 442 C; 502 G; 648 T; 0 other;

Query Match 75.2%; Score 15.8; DB 22; Length 2285;
 Best Local Similarity 89.5%; Pred. No. 1.5e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgctggcggtggaatat 20
 ||||| |||||
 Db 1046 tgctggttggtggaatat 1064

RESULT 12

AAF83002
 ID AAF83002 standard; cDNA; 2437 BP.

XX AAF83002;

XX 29-JUN-2001 (first entry)

DE Human MBSP10 polypeptide encoding cDNA (clone 20604798.0.1).

XX MBSPX; cancer; preclampsia; immune system; neurological; cytostatic;
 KW gynecological; antiinflammatory; neuroprotective; inotropic; relaxant;
 KW cardiant; dermatological; gene therapy; human; MBSP10; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
 FH CDS 147..1597
 FT /*tag= a
 FT /product= "MBSP10"

XX WO200127277-A2.

XX 19-APR-2001.

XX 13-OCT-2000; 2000WO-US28480.

XX (GETH) GENENTECH INC.

XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;

PR 13-OCT-1999; 99US-0159231.
 PR 12-JAN-2000; 2000US-0175670.
 PR 12-OCT-2000; 2000US-0159231.
 XX (CURA-) CURAGEN CORP.
 XX PA
 XX Shimkets RA, Lichenstein H, Boldog FL;
 PI
 XX
 DR WPI; 2001-282030/29.
 DR P-PSDB; AAB62406.
 XX
 XX Novel human polynucleotide sequences and the membrane bound or secreted
 PT polypeptides encoded by these sequences, designated MBSPX -
 XX
 PS Claim 9; Page 44-47; 157pp; English.
 XX
 XX The invention relates to novel polypeptides, termed MBSPX and
 CC polynucleotides encoding the MBSPX polypeptides. The MBSPX polypeptide,
 CC nucleic acid and an MBSPX antibody are useful for treating or preventing
 CC a pathology associated with the protein especially in humans. The MBSPX
 CC nucleic acid can be used to express MBSPX protein (e.g. via a recombinant
 CC expression vector in a host cell in gene therapy applications), an to
 CC detect MBSPX mRNA in a biological sample or a genetic lesion in a MBSPX
 CC gene. Disorders associated with insufficient or excessive production of
 CC MBSPX protein include cancer, preclampsia, immune system disorders and
 CC inflammation, neurological disorders, cardiovascular disorders; and skin
 CC and muscle abnormalities. The anti-MBSPX antibodies can be used to detect
 CC and isolate MBSPX proteins and modulate MBSPX activity. The present
 CC sequence represents the nucleotide sequence of MBSP10.
 XX
 SQ Sequence 2437 BP; 739 A; 473 C; 530 G; 695 T; 0 other;
 Query Match 75.2%; Score 15.8; DB 22; Length 2437;
 Best Local Similarity 89.5%; Pred. No. 1.5e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 tgctggcggtggaatat 20
 Db 1190 tgctgggtggtggaatat 1208
 ||||| |||||
 RESULT 13
 AAF59608
 ID AAF59608 standard; cDNA; 2449 BP.
 XX
 AC AAF59608;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Human cell cycle and proliferation protein CCYPR-19 cDNA, SEQ ID NO:73.
 KW Cell cycle and proliferation protein; CCYPR; human; agonist;
 KW antagonist; gene therapy; detection; gene therapy;
 KW transgenic animal disease model; immune disorder;
 KW developmental disorder; cell signalling disorder;
 KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
 KW arteriosclerosis; asthma; allergy; diabetes mellitus;
 KW menstrual cycle disorder; bacterial infection; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200107471-A2.
 XX
 PD 01-FEB-2001.
 XX
 PF 21-JUL-2000; 2000WO-US19948.
 XX
 PR 21-JUL-1999; 99US-0145075.
 PR 08-SEP-1999; 99US-0153129.
 PR 10-NOV-1999; 99US-0164647.
 XX
 PA (INCY-) INCYTE GENOMICS INC.

XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
 PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
 XX
 DR WPI; 2001-112727/12.
 DR P-PSDB; AAB60471.
 XX
 XX Human cell cycle and proliferation proteins and polynucleotides are
 PT used to treat, diagnose and prevent immune, developmental and cell
 PT signalling disorders and cell proliferative disorders including cancer -
 XX
 XX Example V; Page 179-180; 205pp; English.
 PS
 XX Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human
 CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.
 CC CCYPR and agonists of CCYPR are used to treat diseases or conditions
 CC associated with decreased expression of functional CCYPR, while CCYPR
 CC antagonists are used to treat diseases or conditions associated with
 CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
 CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
 CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
 CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
 CC that specifically bind to CCYPR, and in drug screening methods to
 CC identify compounds that modulate the activity of CCYPR. CCYPR
 CC nucleotides can be used to generate transgenic animal models of human
 CC disease, and can be used in gene therapy in target cells with genetic
 CC abnormalities with respect to the expression of CCYPR for the
 CC treatment or prevention of a disorder associated with CCYPR.
 CC Diseases which can be diagnosed, treated and prevented using CCYPR
 CC proteins, nucleic acids, agonists or antagonists include immune,
 CC developmental and cell signalling disorders, and cell proliferative
 CC disorders including cancer. Specific examples of these disorders
 CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
 CC diabetes mellitus, disorders of the menstrual cycle and infections
 CC caused by bacteria.
 XX
 SQ Sequence 2449 BP; 749 A; 478 C; 531 G; 691 T; 0 other;
 Query Match 75.2%; Score 15.8; DB 22; Length 2449;
 Best Local Similarity 89.5%; Pred. No. 1.5e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 tgctggcggtggaatat 20
 Db 1203 tgctgggtggtggaatat 1221
 ||||| |||||
 RESULT 14
 AAS45041
 ID AAS45041 standard; cDNA; 2495 BP.
 XX
 AC AAS45041;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE cDNA encoding novel human secretory protein, Seq ID No 122.
 XX
 KW Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
 KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
 KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
 KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
 KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;
 KW gut protection; lung; liver fibrosis; immune deficiency; infection;
 KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;
 KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
 KW fertility; analgesic; pain; antigen; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200166689-A2.
 XX
 PD 13-SEP-2001.

XX PF 05-MAR-2001; 2001WO-US04942.
XX PR 07-MAR-2000; 2000US-0519705.
XX PR 19-MAY-2000; 2000US-0574454.
XX PR 17-JUN-2000; 2000US-0596193.
XX PR 14-JUL-2000; 2000US-0616847.
XX PR 19-SEP-2000; 2000US-0665363.
XX PR 20-OCT-2000; 2000US-0693267.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;
XX PI Zhao QA, Yang Y, Brmanac RT, Zhang J, Chen R, Xue AJ, Wang J;
XX WPI: 2001-589934/66.
XX DR P-PSDB; AAU28141.
XX XX Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis and treatment of
PT cancer, neurological, inflammatory, and autoimmune disorders -
XX
XX Claim 1; SEQ ID No 122; 107pp; English.
XX
XX The invention relates to novel isolated human secreted polypeptides (I)
CC and polynucleotides (II). (I) and (II) are useful for treating
CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,
CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is
CC involved in increasing haematopoiesis, stem cell survival, bone growth
CC and remodeling. (I), (II) and modulators of (II) are useful for
CC prophylaxis or treatment of one or more cancers. (II) is also useful for
CC creating transgenic animals useful for studying the in vivo activities of
CC the polypeptide as well as for studying modulators of the polypeptides.
CC (I) induces the proliferation of neural cells and regeneration of nerve
CC and brain tissue and is useful for the treatment of central and
CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,
CC Parkinson's disease, Huntington's disease, and amyotrophic lateral
CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth, and in tissue repair, healing of burns, incisions,
CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative
CC disorders, or periodontal disease. Furthermore, (I) is also useful for
CC gut protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, various immune deficiencies and
CC disorders including severe combined immunodeficiency (SCID), bacterial or
CC fungal infections, autoimmune disorders e.g. multiple sclerosis,
CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
CC reactions and conditions, such as asthma or other respiratory problems.
CC In addition, (I) affects biorhythms or circadian cycles of rhythms,
CC fertility, metabolism, catabolism, anabolism, storage or elimination of
CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
CC analgesic effects or other pain reducing effects, immunoglobulin like
CC activity and can act as an antigen in a vaccine composition to raise an
CC immune response. AAS44920-AAS45295 represent novel human secreted protein
CC coding sequences of the invention.
XX
XX Sequence 2495 BP; 759 A; 484 C; 553 G; 699 T; 0 other;
SQ

Query Match 75.2%; Score 15.8; DB 22; Length 2495;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 tgcggcggtggaatat 20
||||| |||||
Db 1236 tgcgggttggtggaatat 1254

RESULT 15
AAC76426
ID AAC76426 standard; cDNA; 2505 BP.

XX AC AAC76426;
XX DT 08-FEB-2001 (first entry)
XX DE Human ORFX ORF1981 polynucleotide sequence SEQ ID NO:3961.
XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antihydroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX OS Homo sapiens.
XX PN WO200058473-A2.
XX PD 05-OCT-2000.
XX PF 31-MAR-2000; 2000WO-US08621.
XX PR 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX PA (CURA-) CURAGEN CORP.
XX PI Shinkets RA, Leach M;
XX WPI: 2000-602362/57.
XX DR P-PSDB; AAB42217.
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX Claim 5; Page 3123-3124; 5507pp; English.
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antihydroid; and antianaemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy.
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX Sequence 2505 BP; 774 A; 482 C; 547 G; 700 T; 2 other;

Query Match 75.2%; Score 15.8; DB 21; Length 2505;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgtcgggcgtggaatat 20
||||| |||||||||
Db 1215 tgtcgttggtggaatat 1233

Search completed: July 25, 2002, 05:44:43
Job time: 6951 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 05:38:29 ; Search time 90.93 Seconds
(without alignments)
56.728 Million cell updates/sec

Title: US-09-235-416-3
Perfect score: 21
Sequence: 1 atgtcggcggtggaatc 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	15.2	72.4	60	3	US-09-023-228B-91
C 2	15.2	72.4	61	3	US-09-023-228B-131
C 3	15.2	72.4	448	4	US-09-060-756-666
C 4	15.2	72.4	2424	1	US-08-821-119-16
C 5	15.2	72.4	3350	3	US-08-617-8608-3
6	15	71.4	1533	1	US-07-721-761A-32
7	15	71.4	1533	1	US-07-978-687-32
8	15	71.4	1533	5	US-08-471-791-12
9	15	71.4	1533	5	PCT-US91-01746-12
10	15	71.4	1533	5	PCT-US91-05801-32
11	14.8	70.5	113	1	US-08-616-368A-23
12	14.8	70.5	113	3	US-09-054-298-23
13	14.8	70.5	113	4	US-08-818-655-23
14	14.8	70.5	840	1	US-08-616-368A-14
15	14.8	70.5	840	3	US-09-054-298-14
16	14.8	70.5	840	4	US-08-818-655-14
17	14.6	69.5	1052	6	5489533-1
18	14.6	69.5	1052	6	5512660-1
C 19	14.2	67.6	769	4	US-08-776-971-118
C 20	14.2	67.6	769	4	US-08-776-971-119
C 21	14.2	67.6	1193	4	US-09-347-798-1
C 22	14.2	67.6	1668	2	US-08-505-377-2
23	14.2	67.6	1668	2	US-08-505-377-5
24	14.2	67.6	1668	3	US-08-798-269-2
25	14.2	67.6	1668	3	US-08-798-269-5
26	14.2	67.6	1668	4	US-09-055-210-2
27	14.2	67.6	1668	4	US-09-055-210-5

28	14.2	67.6	1952	3	US-08-714-918-39	Sequence 39, Appl
29	14.2	67.6	1952	4	US-09-265-315-39	Sequence 39, Appl
30	14.2	67.6	1952	4	US-09-265-315-39	Sequence 39, Appl
31	14.2	67.6	1952	4	US-09-266-417-39	Sequence 39, Appl
C 32	14.2	67.6	2341	3	US-08-714-918-102	Sequence 102, App
C 33	14.2	67.6	2341	4	US-09-265-315-102	Sequence 102, App
C 34	14.2	67.6	2341	4	US-09-265-315-102	Sequence 102, App
C 35	14.2	67.6	2341	4	US-09-266-417-102	Sequence 102, App
36	14	66.7	935	3	US-09-187-331-3	Sequence 3, Appl1
37	14	66.7	935	4	US-09-470-946-3	Sequence 267, App
38	13.8	65.7	49	1	US-08-171-389-267	Sequence 267, App
39	13.8	65.7	49	1	US-08-123-936-267	Sequence 267, App
40	13.8	65.7	49	2	US-08-475-228A-267	Sequence 267, App
41	13.8	65.7	49	3	US-08-482-080A-267	Sequence 267, App
42	13.8	65.7	49	5	PCT-US93-12388-267	Sequence 267, App
C 43	13.8	65.7	827	4	US-08-976-259-36	Sequence 36, Appl
C 44	13.8	65.7	840	4	US-08-998-416-507	Sequence 507, App
45	13.8	65.7	1303	2	US-08-793-410-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-023-228B-91/c
; Sequence 91, Application US/09023228B
; Patent No. 6140490
; GENERAL INFORMATION:
; APPLICANT: BIESECKER, GREGORY
; APPLICANT: GOLD, LARRY
; TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF
; NUMBER OF SEQUENCES: 157
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 E. Prentice Place #200
; CITY: Denver
; STATE: Colorado
; COUNTRY: US
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,228B
; FILING DATE: 12-FEBRUARY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: PCT/US97/01739
; APPLICATION NUMBER: 30 JAN 1997
; FILING DATE: 30 JAN 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/595,335
; FILING DATE: 1 FEB 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson, Esq.
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX50/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA
; FEATURE:
; OTHER INFORMATION: All pyrimidines are 2'-F modified
US-09-023-228B-91

US-08-821-119-16

Query Match 72.4%; Score 15.2; DB 1; Length 2424;
Best Local Similarity 85.0%; Pred. No. 41;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tctcggcggtggaataatc 21
|||||
Db 463 TGTCGGGGGTGCACATATC 444

RESULT 5

US-08-617-860B-3/c
; Sequence 3, Application US/08617860B
; Patent No. 6133506
; GENERAL INFORMATION:
; APPLICANT: Typfer, R., Bautor, J., Bothmann, H., Filsak, E.,
; APPLICANT: Hvircke-Grandpierre, C., Klein, B., Martini, N.,
; APPLICANT: Mller, A., Schulte, W., Voetz, M., Walek, J.,
; APPLICANT: Schell, J.
; TITLE OF INVENTION: Promoters
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
; STREET: 1140 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,860B
; FILING DATE: 01-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/02950
; FILING DATE: 05-SEP-1994
; APPLICATION NUMBER: DE P4329951.2
; FILING DATE: 04-SEP-1993
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3350 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Double stranded
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Brassica napus
; IMMEDIATE SOURCE:
; LIBRARY: genomic Lambda FIX II
; CLONE: BnAccaseg10
; FEATURE:
; NAME/KEY: Startcodon
; LOCATION: 2611..2613
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(2611..2908, 3001..3341)
US-08-617-860B-3

Query Match 72.4%; Score 15.2; DB 3; Length 3350;
Best Local Similarity 85.0%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atctcggcggtggaataat 20
|||||
Db 600 ATGGCGGACGGTGGTAATAT 581

RESULT 6

US-07-721-761A-32
; Sequence 32, Application US/07721761A
; Patent No. 5475099
; GENERAL INFORMATION:
; APPLICANT: Vic. C. Knauf
; APPLICANT: Gregory A. Thompson
; TITLE OF INVENTION: Plant Fatty Acid Synthases
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.7
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/721,761A
; FILING DATE: 19910626
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/568,493
; FILING DATE: 15-AUGUST-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; REFERENCE/DOCKET NUMBER: CGNE 76-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1533 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
US-07-721-761A-32

Query Match 71.4%; Score 15; DB 1; Length 1533;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gtcggcggtggaataat 17
|||||
Db 1117 GTCCGGCGGTGGAA 1131

RESULT 7

US-07-978-687-32
; Sequence 32, Application US/07978687
; Patent No. 5510255
; GENERAL INFORMATION:
; APPLICANT: Vic. C. Knauf
; APPLICANT: Gregory A. Thompson
; TITLE OF INVENTION: Plant Fatty Acid Synthases
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA

```

; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.7
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/978,687
; FILING DATE: FEBRUARY 1, 1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/05801
; FILING DATE: 15-AUGUST-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/568,493
; FILING DATE: 15-AUGUST-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/721,761
; FILING DATE: 26-JUNE-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; REFERENCE/DOCKET NUMBER: CGNE 76-2 WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1533 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-07-978-687-32

```

```

Query Match 71.4%; Score 15; DB 1; Length 1533;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 3 gtcggcggtggaaa 17
Db 1117 GTCGGCGGTGGAAA 1131

```

```

RESULT 8
US-08-471-791-12
; Sequence 12, Application US/08471791
; Patent No. 5723595
; GENERAL INFORMATION:
; APPLICANT: Thompson, Gregory A
; APPLICANT: Knauf, Vic C
; TITLE OF INVENTION: Plant Desaturases-Compositions and Uses
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: California
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.1
; SOFTWARE: Microsoft Word 5.1 (a)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,791
; FILING DATE: 6-JUNE-95
; CLASSIFICATION: 435

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/762,762
; FILING DATE: 16-SEPT-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/01746
; FILING DATE: 14-MAR-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/615,784
; FILING DATE: 14-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/567,373
; FILING DATE: 13-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/494,106
; FILING DATE: 16-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Lassen, Elizabeth
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 69-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; TELEX: 350370 CGNE
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1533 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-08-471-791-12

```

```

Query Match 71.4%; Score 15; DB 1; Length 1533;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 3 gtcggcggtggaaa 17
Db 1117 GTCGGCGGTGGAAA 1131

```

```

RESULT 9
PCT-US91-01746-12
; Sequence 12, Application PC/TUS9101746
; GENERAL INFORMATION:
; APPLICANT: Thompson, Gregory A
; APPLICANT: Knauf, Vic C
; TITLE OF INVENTION: Plant Desaturases-Compositions and Uses
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: California
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.7
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/01746
; FILING DATE: 19910314
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```


APPLICATION NUMBER: 07/615,784
FILING DATE: 14-NOV-1990
APPLICATION NUMBER: 07/567,373
FILING DATE: 13-AUG-1990
APPLICATION NUMBER: 07/494,106
FILING DATE: 16-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Lassen, Elizabeth
REGISTRATION NUMBER: 31,845
NAME: Donna E. Scherer
REGISTRATION NUMBER: 34,719
REFERENCE/DOCKET NUMBER: CGNE 69-3 WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (916) 753-6313
TELEFAX: (916) 753-1510
TELEX: 350370 CGNE
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1533 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna to mRNA
PCT-US91-01746-12

Query Match 71.4%; Score 15; DB 5; Length 1533;

Best Local Similarity 100.0%; Pred. No. 50;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gtcgggcggtgga 17

|||||

Db 1117 GTCGGCGGTGGA 1131

RESULT 10
PCT-US91-05801-32
Sequence 32, Application PCT/US9105801
GENERAL INFORMATION:
APPLICANT: Vic. C. Knauf
APPLICANT: Gregory A. Thompson
TITLE OF INVENTION: Plant Fatty Acid Synthases
NUMBER OF SEQUENCES:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Calgene, Inc.
STREET: 1920 Fifth Street
CITY: Davis
STATE: CA
COUNTRY: USA
ZIP: 95616
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.7
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/05801
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/568,493
FILING DATE: 15-AUGUST-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/721,761
FILING DATE: 26-JUNE-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elizabeth Lassen
REGISTRATION NUMBER: 31,845
NAME: Donna E. Scherer
REGISTRATION NUMBER: 34,719
REFERENCE/DOCKET NUMBER: CGNE 76-2 WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (916) 753-6313

TELEFAX: (916) 753-1510
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 1533 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna to mRNA
PCT-US91-05801-32

Query Match 71.4%; Score 15; DB 5; Length 1533;

Best Local Similarity 100.0%; Pred. No. 50;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gtcgggcggtgga 17

|||||

Db 1117 GTCGGCGGTGGA 1131

RESULT 11
US-08-616-368A-23
Sequence 23, Application US/08616368A
Patent No. 5767262
GENERAL INFORMATION:
APPLICANT: Lee, Mu-En
APPLICANT: Haber, Edgar
APPLICANT: Jain, Mukesh
APPLICANT: Yet, Shaw-Fang
TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/616,368A
FILING DATE: 15-MAR-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/022001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 113 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-616-368A-23

Query Match 70.5%; Score 14.8; DB 1; Length 113;

Best Local Similarity 88.9%; Pred. No. 45;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcgggcggtgga 19

|||||

Db 9 TCTGGGCGGTGGAATA 26

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 04:39:04 ; Search time 2969.55 Seconds
(without alignments)
95.447 Million cell updates/sec

Title: US-09-235-416-3
Perfect score: 21
Sequence: 1 atgtcggcggtggaataatc 21

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
1: em_estba:*
2: em_esthum:*
3: em_estinu:*
4: em_estinu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.8	80.0	288	9	BB607934
2	16.8	80.0	428	9	AV213384
3	16.8	80.0	460	9	AW829820
4	16.8	80.0	613	12	BH532588
5	16.8	80.0	568	10	BG160433
6	16.8	80.0	792	10	B5627929
7	16.8	80.0	1278	12	B09467
8	16.8	80.0	1293	10	BH415118
9	16.4	78.1	1406	10	B1666361
10	16.2	77.1	161	12	BH205870
11	16.2	77.1	161	12	BH209290
12	16.2	77.1	161	12	BH209293
13	16.2	77.1	373	10	BG362171
14	16.2	77.1	423	10	BG409025
15	16.2	77.1	463	9	AW473489
16	16.2	77.1	506	10	BF082179
17	16.2	77.1	555	10	B1880600

18	16.2	77.1	598	10	B1891609
19	16.2	77.1	614	12	BH378138
20	16.2	77.1	666	12	AG150846
21	16.2	77.1	719	12	AG163162
22	16.2	77.1	731	10	BE258413
23	16.2	77.1	801	9	BE040997
24	16.2	77.1	822	12	BH026906
25	16.2	77.1	918	10	BE964839
26	16.2	77.1	955	10	BE966691
27	16.2	77.1	1101	12	CNS00200
28	16.2	77.1	2232	10	BF128479
29	16.2	77.1	834	12	AZ186641
30	15.8	75.2	277	9	BB108759
31	15.8	75.2	287	9	BB010176
32	15.8	75.2	289	9	BB598446
33	15.8	75.2	294	9	AA322509
34	15.8	75.2	294	9	BB126311
35	15.8	75.2	295	9	AA380119
36	15.8	75.2	297	10	N48393
37	15.8	75.2	304	10	BF377898
38	15.8	75.2	305	10	BF377896
39	15.8	75.2	316	10	W20331
40	15.8	75.2	317	9	AW799635
41	15.8	75.2	338	10	Z44602
42	15.8	75.2	349	10	W53617
43	15.8	75.2	357	9	AI157155
44	15.8	75.2	417	10	N25409
45	15.8	75.2	418	9	AI120786

ALIGNMENTS

RESULT 1

LOCUS	BB607934	288 bp	mRNA	linear	EST 06-DEC-2000
DEFINITION	BB607934	RIKEN full-length enriched, 2 days pregnant adult female oviduct Mus musculus cDNA clone E230008G20 5', mRNA sequence.			
ACCESSION	BB607934	1	GI:11563111		
VERSION	BB607934	1	GI:11563111		
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	1	(bases 1 to 288)			
AUTHORS	Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.				
TITLE	RIKEN Mouse ESTs (Aizawa,K. et al. 2000)				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome.res@gs.c.riken.go.jp URL: http://genome.gsc.riken.go.jp/ Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Thermolabile and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) Itoh,M., Kitsuunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki				

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenro-cho, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki,
Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh, M., Kitsumi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES

source

Location/Qualifiers

1. 428

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="2410127F20"

/clone_lib="RIKEN full-length enriched, ES cells"

/cell_type="ES cells"

/lab_host="SOLR"

/note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTTCTTTTCTTTTNN 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FIC 1."

BASE COUNT 53 a 67 c 111 g 57 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 9; Length 288;

Best Local Similarity 90.0%; Pred. No. 8.1e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcggcggtgggaataac 21

Db 31 TGTCTGGCGGTGGGAATGTC 50

RESULT 2

AV213384/c

LOCUS

DEFINITION

AV213384 RIKEN full-length enriched, ES cells Mus musculus cDNA clone 2410127F20 3' similar to AF050165 Mus musculus protein phosphatase 2A regulatory subunit PR59 mRNA, mRNA sequence.

ACCESSION

AV213384.1 GI:6152886

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 428)

REFERENCE

AUTHORS

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al. 1999)

Unpublished (1999)

TITLE

JOURNAL

Query Match 80.0%; Score 16.8; DB 9; Length 428;

Best Local Similarity 90.0%; Pred. No. 8.7e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcggcggtgggaataac 20

Db 86 ATGTCGGCGGTGGGAATAT 67

RESULT 3

AW829820

LOCUS

DEFINITION

AW829820 RIKEN full-length enriched, ES cells Mus musculus cDNA clone 2410127F20 3' similar to AF050165 Mus musculus protein phosphatase 2A regulatory subunit PR59 mRNA, mRNA sequence.

ACCESSION

AW829820.1 GI:7923632

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

southern root-knot nematode.

Meloidogyne incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchida; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

1 (bases 1 to 460)

REFERENCE

AUTHORS

ra44g10.y1 Bird-Rao Meloidogyne incognita j2 Meloidogyne incognita

CDNA 5', mRNA sequence.

AW829820

DEFINITION

AW829820 RIKEN full-length enriched, ES cells Mus musculus cDNA clone 2410127F20 3' similar to AF050165 Mus musculus protein phosphatase 2A regulatory subunit PR59 mRNA, mRNA sequence.

ACCESSION

AW829820.1 GI:7923632

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

southern root-knot nematode.

Meloidogyne incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchida; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

1 (bases 1 to 460)

REFERENCE

AUTHORS

AW829820

LOCUS

DEFINITION

AW829820 RIKEN full-length enriched, ES cells Mus musculus cDNA clone 2410127F20 3' similar to AF050165 Mus musculus protein phosphatase 2A regulatory subunit PR59 mRNA, mRNA sequence.

ACCESSION

AW829820.1 GI:7923632

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

southern root-knot nematode.

Meloidogyne incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchida; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

1 (bases 1 to 460)

REFERENCE

AUTHORS

AW829820

LOCUS

DEFINITION

AW829820 RIKEN full-length enriched, ES cells Mus musculus cDNA clone 2410127F20 3' similar to AF050165 Mus musculus protein phosphatase 2A regulatory subunit PR59 mRNA, mRNA sequence.

ACCESSION

AW829820.1 GI:7923632

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

southern root-knot nematode.

Meloidogyne incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchida; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

1 (bases 1 to 460)

REFERENCE

AUTHORS

AW829820

LOCUS

DEFINITION

AW829820 RIKEN full-length enriched, ES cells Mus musculus cDNA clone 2410127F20 3' similar to AF050165 Mus musculus protein phosphatase 2A regulatory subunit PR59 mRNA, mRNA sequence.

ACCESSION

AW829820.1 GI:7923632

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

southern root-knot nematode.

Meloidogyne incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchida; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

1 (bases 1 to 460)

REFERENCE

AUTHORS

AW829820

LOCUS

DEFINITION

AW829820 RIKEN full-length enriched, ES cells Mus musculus cDNA clone 2410127F20 3' similar to AF050165 Mus musculus protein phosphatase 2A regulatory subunit PR59 mRNA, mRNA sequence.

ACCESSION

AW829820.1 GI:7923632

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

southern root-knot nematode.

Meloidogyne incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchida; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

1 (bases 1 to 460)

REFERENCE

AUTHORS

AW829820

LOCUS

DEFINITION

AW829820 RIKEN full-length enriched, ES cells Mus musculus cDNA clone 2410127F20 3' similar to AF050165 Mus musculus protein phosphatase 2A regulatory subunit PR59 mRNA, mRNA sequence.

ACCESSION

AW829820.1 GI:7923632

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

southern root-knot nematode.

Meloidogyne incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchida; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

1 (bases 1 to 460)

REFERENCE

AUTHORS

AW829820

LOCUS

DEFINITION

AW829820 RIKEN full-length enriched, ES cells Mus musculus cDNA clone 2410127F20 3' similar to AF050165 Mus musculus protein phosphatase 2A regulatory subunit PR59 mRNA, mRNA sequence.

ACCESSION

AW829820.1 GI:7923632

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

southern root-knot nematode.

Meloidogyne incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchida; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

1 (bases 1 to 460)

REFERENCE

AUTHORS

AW829820

LOCUS

DEFINITION

AW829820 RIKEN full-length enriched, ES cells Mus musculus cDNA clone 2410127F20 3' similar to AF050165 Mus musculus protein phosphatase 2A regulatory subunit PR59 mRNA, mRNA sequence.

ACCESSION

AW829820.1 GI:7923632

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

southern root-knot nematode.

Meloidogyne incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchida; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

1 (bases 1 to 460)

REFERENCE

AUTHORS

AW829820

LOCUS

DEFINITION

AW829820 RIKEN full-length enriched, ES cells Mus musculus cDNA clone 2410127F20 3' similar to AF050165 Mus musculus protein phosphatase 2A regulatory subunit PR59 mRNA, mRNA sequence.

ACCESSION

AW829820.1 GI:7923632

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

southern root-knot nematode.

Meloidogyne incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchida; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

1 (bases 1 to 460)

REFERENCE

AUTHORS

AW829820

LOCUS

DEFINITION

AW829820 RIKEN full-length enriched, ES cells Mus musculus cDNA clone 2410127F20 3' similar to AF050165 Mus musculus protein phosphatase 2A regulatory subunit PR59 mRNA, mRNA sequence.

ACCESSION

AW829820.1 GI:7923632

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

southern root-knot nematode.

Meloidogyne incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchida; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

1 (bases 1 to 460)

REFERENCE

AUTHORS</

AUTHORS

McCarter, J., Clifton, S., Chlapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelis, W., R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)

JOURNAL

Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

The library was constructed by Uma Rao and David Bird

(david.bird@ncsu.edu) at North Carolina State University. DNA

Sequencing by: Washington University Genome Sequencing Center St.

Louis.

Seq primer: -40RP from Gibco

High quality sequence stop: 438.

Location/Qualifiers

FEATURES

source

1..460

/organism="Meloiodogyne incognita"

/db_xref="taxon:6306"

/clone_lib="Bird-Rao Meloiodogyne incognita J2"

/dev_stage="enriched for 2nd stage juveniles"

/lab_host="XLORL"

/note="Vector: ZAP express - pBKCWV (Stratagene); Site_1:

EcoRI; Site_2: XhoI; Oligo (dT) primed library. cDNA was

constructed and cloned unidirectionally into the vector

within the 5' EcoRI and 3' XhoI sites. This library was

constructed by Dr. Uma Rao and Dr. David Bird at North

Carolina state University."

160 a 80 c 86 g 134 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 80.0%; Score 16.8; DB 9; Length 460;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgtcggcggtggaatc 21

||||| ||| ||||| |||||

Db 198 TGTGAGGAGTGGAAATATC 217

RESULT 4

BH532588/c

LOCUS

DEFINITION BH532588 BOGH Brassica oleracea genomic clone BOGH87, DNA

sequence.

ACCESSION BH532588

VERSION BH532588.1 GI:17754954

KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 613)

Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other_GSSs: BOGH87TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

FEATURES

source

Location/Qualifiers

1..613

/organism="Brassica oleracea"

/strain="TO1000DH3"

/db_xref="taxon:3712"

/clone="BOGH87"

/clone_lib="BOGH"

/note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared

genomic DNA inserted into PHOS1 using BstXI linkers"

176 a 113 c 106 g 218 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 80.0%; Score 16.8; DB 12; Length 613;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgcggcggtggaatat 20

||||| ||| ||||| |||||

Db 266 ATGTCGGTGTGAAATAT 247

RESULT 5

BG160433/c

LOCUS

DEFINITION BG160433 668 bp mRNA linear EST 06-FEB-2001

IMAGE:3580245 5' similar to SW:CRS3_HORSE O19010 CYSTEINE-RICH

SECRETORY PROTEIN-3 PRECURSOR. [1]; mRNA sequence.

ACCESSION BG160433

VERSION BG160433.1 GI:12694352

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 668)

Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.

, Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person

, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,

Waterston, R. and Wilson, R.

WashU Xenopus EST project, 1999

Unpublished (1999)

Contact: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library constructed by N. Garrett, K. Ryan and A.M. Zorn,

(Wellcome/CRC Institute). DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov

Seq primer: -40RP from Gibco

High quality sequence stop: 513.

FEATURES

source

Location/Qualifiers

1..668

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="IMAGE:3580245"

/clone_lib="Wellcome CRC PRN3 head"

/tissue_type="head, stage 30"

/lab_host="DH10B (phage-resistant)"

/note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs

were oligo-dT primed and directionally cloned. Staging

according to Nieuwkoop and Faber. Library was constructed

by N. Garrett, K. Ryan and A.M. Zorn, (Wellcome/CRC

Institute)."

197 a 162 c 143 g 165 t 1 others

BASE COUNT

Db 1094 TGTGGGGCGGTGATATATC 1113

RESULT 8

BM415118 1293 bp mRNA linear EST 28-JAN-2002
 LOCUS OP20190 Mixed Stage EST's from Globodera pallida, the potato cyst
 DEFINITION nematode Globodera pallida cDNA, mRNA sequence.

BM415118

ACCESSION

BM415118

VERSION

BM415118.1 GI:18381473

KEYWORDS

EST.

SOURCE

ORGANISM Globodera pallida.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
 1 (bases 1 to 1293)
 AUTHORS Heer, J., Sosinski, B., Pokrzywa, R.M., Warry, A. and Opperman, C.
 TITLE Mixed Stage EST's from Globodera pallida, the potato cyst nematode
 JOURNAL Unpublished (2001)
 COMMENT Contact: Opperman, C
 Center for the Biology of Nematode Parasitism
 NC State University; IACR-Rothamsted
 Campus Box 7616; Raleigh, NC 27695, USA
 Tel: 919.515.6699
 Fax: 919.515.9500
 Email: warthog@unity.ncsu.edu
 GT11-11PCN_R_H11_GT11-11_R_092.ab1.

FEATURES

source

1..1293

/organism="Globodera pallida"
 /db_xref="taxon:36090"
 /clone_lib="Mixed Stage EST's from Globodera pallida, the
 potato cyst nematode"
 /note="Vector: lambda GT11; This is a collaborative effort
 between IACR-Rothamsted and North Carolina State
 University. The library was constructed from mixed stage
 G. pallida in lambda GT11 by Paul Burroughs,
 IACR-Rothamsted."

BASE COUNT

ORIGIN 327 a 340 c 282 g 320 t 24 others

Query Match

Best Local Similarity 80.0%; Score 16.8; DB 10; Length 1293;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaatat 20

|||||||

Db 115 ATGTCGGGCGCTAGAAATAT 134

RESULT 9

BI666361/c

LOCUS

DEFINITION

BI666361 1406 bp mRNA linear EST 12-SEP-2001
 603290866F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5325093 5',
 mRNA sequence.

ACCESSION

BI666361

VERSION

BI666361.1 GI:15580594

KEYWORDS

EST.

SOURCE

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1406)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LLAM1824 row: 1 column: 22

High quality sequence start: 18

High quality sequence stop: 75.

FEATURES

source

1..1406

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone_lib="IMAGE:5325093"

/sex="female, virgin"

/tissue_type="infiltrating ductal carcinoma"

/dev_stage="5 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Library constructed by Life Technologies. Investigator
 providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 393 a 338 c 296 g 379 t

ORIGIN

Query Match 78.1%; Score 16.4; DB 10; Length 1406;

Best Local Similarity 94.4%; Pred. No. 1.7e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 gtctggcggtggaatat 20

|||||||

Db 1184 GACGGGCGGTGGAATAT 1167

RESULT 10

BH205870

LOCUS

DEFINITION

BH205870 161 bp DNA linear GSS 24-OCT-2001
 Sml-54D8.TR Sml Schistosoma mansoni genomic clone Sml-54D8, DNA
 sequence.

ACCESSION

BH205870

VERSION

BH205870.1 GI:16381631

KEYWORDS

GSS.

SOURCE

ORGANISM Schistosoma mansoni.

Schistosoma mansoni

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.

REFERENCE 1 (bases 1 to 161)

AUTHORS Shetty, J., Simpson, A., Malek, J., Koo, H., LoVerde, P.T. and El-Sayed

, N.M.

TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain)

Sml BAC library for gene discovery and map construction

Unpublished (2001)

Other_GSSs: Sml-54D8.TF

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: nelsayed@tigr.org

10.edu).

Seq primer: M13 Rev

Class: BAC ends.

Location/Qualifiers

1..161

/organism="Schistosoma mansoni"

/strain="Puerto Rico"

/db_xref="taxon:6183"

/clone_lib="Sml-54D8"

/clone_lib="Sml"

/note="Vector: pBelOAC11; Site_1: Hin dIII; Constructed

in the laboratory of Dr. Denis Le Paslier at the Fondation

Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma

mansonii agarose embedded DNA was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 7.95 x the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."

44 a 32 c 41 g 44 t

BASE COUNT
ORIGIN

Query Match 77.1%; Score 16.2; DB 12; Length 161;
Best Local Similarity 85.7%; Pred. No. 1.4e+03;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgtcggcggtggaataatc 21
||||| ||||| ||||| ||

Db 16 ATGTCGGTCGTTGAAATGTC 36

RESULT 11
BH209290/c

LOCUS Sml-56Pl.TF Sml Schistosoma mansoni genomic clone Sml-56Pl, DNA
DEFINITION sequence.

ACCESSION BH209290

VERSION BH209290.1 GI:16387816

KEYWORDS GSS.

SOURCE Schistosoma mansoni.

ORGANISM

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.

1 (bases 1 to 161)

REFERENCE Shetty,J., Simpson,A., Malek,J., Koo,H., Loverde,P.T. and El-Sayed

AUTHORS ,N.M.

TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain)

Sml BAC library for gene discovery and map construction

Unpublished (2001)

Other_GSSs: Sml-56Pl.TF

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: nelsayed@tigr.org

Seq primer: M13 For

Class: BAC ends.

Location/Qualifiers

1..161

/organism="Schistosoma mansoni"

/strain="Puerto Rico"

/db_xref="taxon:6183"

/clone="Sml-56Pl"

/clone_lib="Sml"

/note="Vector: pBelBAC11; Site_1: Hin dIII; Constructed in the laboratory of Dr. Denis Le Paslier at the Fondation Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma mansoni agarose embedded DNA was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 7.95 x the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."

44 a 41 c 32 g 44 t

BASE COUNT
ORIGIN

Query Match 77.1%; Score 16.2; DB 12; Length 161;

Best Local Similarity 85.7%; Pred. No. 1.4e+03;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgtcggcggtggaataatc 21

||||| ||||| ||||| ||

Db 146 ATGTCGGTCGTTGAAATGTC 126

RESULT 12

BH209293

LOCUS Sml-56Pl.TF Sml Schistosoma mansoni genomic clone Sml-56Pl, DNA

DEFINITION sequence.

ACCESSION BH209293

VERSION BH209293.1 GI:16387822

KEYWORDS GSS.

SOURCE Schistosoma mansoni.

ORGANISM

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.

1 (bases 1 to 161)

REFERENCE Shetty,J., Simpson,A., Malek,J., Koo,H., Loverde,P.T. and El-Sayed

AUTHORS ,N.M.

TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain)

Sml BAC library for gene discovery and map construction

Unpublished (2001)

Other_GSSs: Sml-56Pl.TF

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: nelsayed@tigr.org

Seq primer: M13 Rev

Class: BAC ends.

Location/Qualifiers

1..161

/organism="Schistosoma mansoni"

/strain="Puerto Rico"

/db_xref="taxon:6183"

/clone="Sml-56Pl"

/clone_lib="Sml"

/note="Vector: pBelBAC11; Site_1: Hin dIII; Constructed in the laboratory of Dr. Denis Le Paslier at the Fondation Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma mansoni agarose embedded DNA was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 7.95 x the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."

44 a 32 c 41 g 44 t

BASE COUNT
ORIGIN

Query Match 77.1%; Score 16.2; DB 12; Length 161;

Best Local Similarity 85.7%; Pred. No. 1.4e+03;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgtcggcggtggaataatc 21

||||| ||||| ||||| ||

Db 16 ATGTCGGTCGTTGAAATGTC 36

RESULT 13

BG362171

LOCUS BG362171

DEFINITION 9B52f07.y1 Moss EST library PPG Physcomitrella patens cDNA clone

PEP_SOURCE_ID: 5', mRNA sequence.

373 bp mRNA linear EST 08-MAR-2001

```

ACCESSION   BG362171
VERSION     BG362171.1  GI:13251268
KEYWORDS    EST.
SOURCE      Physcomitrella patens.
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1 (bases 1 to 373)
AUTHORS     Quatrano, R., Bashlades, S., Cove, D., Cumling, A., Knight, C., Clifton
            , S., Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood
            , K., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T.,
            Steptoe, M., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.
            , Waterston, R. and Wilson, R.
TITLE       Leeds/Wash U Moss EST Project
JOURNAL     Unpublished (1999)
COMMENT     Contact: Ralph Quatrano
            Leeds/Wash U Moss EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Libraries were constructed by Dr. Stavros Bashlades as part of the
            Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
            Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
            University Genome Sequencing Center For information on obtaining a
            clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 338.

FEATURES    Location/Qualifiers
            1..373
             /organism="Physcomitrella patens"
             /db_xref="taxon:3218"
             /clone="PEP_SOURCE ID:"
             /clone_lib="Moss EST library PPG"
             /tissue_type="gametophore: 30 day old tissue,
             ammonium-grown"
             /lab_host="DH10B"
             /note="Vector: pAMP1; Construction of the cDNA library was
             performed by Dr. W. Gregg Clark using a modification of
             the cDNA synthesis protocol developed in the laboratory of
             Dr. Michael Lovett by Dr. Yulia Korshunova (personal
             communication). First polyA + RNA was isolated from total
             gametophore RNA using oligo dt magnetic beads. Following
             this, first strand cDNA synthesis was performed on the
             bead-bound polyA + RNA, during which an oligonucleotide
             anchor sequence was incorporated onto the 5'-ends of the
             cDNA. PCR amplification was then used to synthesize the
             second strand, to amplify the double stranded DNA, and to
             incorporate dUTP containing sequences into the ends of the
             double stranded cDNA. This DNA was size selected and
             cloned into pAMP1 using the CloneAMP pAMP1 System (Life
             Technologies, GibcoBRL) for cloning amplification products
             by a non-restriction site dependant process. The cloning
             was directional based on sequence asymmetry introduced at
             the ends during PCR amplification. The 3' cDNA ends are
             proximal to the NotI site of the multiple cloning site in
             pAMP1. This annealing mixture was transformed into
             chemically competent DH10B cells and selected for
             ampicillin resistant growth. The resulting clones (about
             330,000) were pooled to make the library."
BASE COUNT   100 a 81 c 86 g 106 t
ORIGIN

Query Match      77.1%; Score 16.2; DB 10; Length 373;
Best Local Similarity 85.7%; Pred. No. 1.6e+03;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atgtcggcggtggaataatc 21
    ||||| ||| ||||| |||
DB 21 ATGCTCTGGCTGTGGAACATC 41

```

Query Match

77.1%; Score 16.2; DB 10; Length 423;

Best Local Similarity 85.7%; Pred. No. 1.7e+03;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgtcggcggtggaataac 21
||||| ||| ||||| |||
Db 86 ATGTCGGCTGTGGAACATC 106

RESULT 15

AW473489 463 bp mRNA linear EST 24-FEB-2000
LOCUS ha75f01.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2879545 3'
DEFINITION similar to contains Alu repetitive element; contains element OFR
repetitive element ; mRNA sequence.

ACCESSION AW473489
VERSION AW473489.1 GI:7043595
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 463)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncilogap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone Distribution: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -400p from Gibco
High quality sequence stop: 401.

FEATURES
source

1..463
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2879545"
/clone_lib="NCI_CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signed ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site:1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"

BASE COUNT 119 a 121 c 91 g 132 t
ORIGIN

Query Match 77.1%; Score 16.2; DB 9; Length 463;
Best Local Similarity 85.7%; Pred. NO. 1.7e+03;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgtcggcggtggaataac 21
||||| ||| ||||| |||
Db 196 ATGTCGGGTGGGGGAATATC 216

Search completed: July 25, 2002, 04:39:12
Job time: 3150 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 05:38:07 ; Search time 3312.9 Seconds
(without alignments)
145.284 Million cell updates/sec

Title: US-09-235-416-4
Perfect score: 23
Sequence: 1 gaattctgcttcgtgttttca 23

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
--------	-------	-------	-------	--------	----	-------------

	1	18.8	81.7	72241	2	AC069180	Homo sapi
	2	18.8	81.7	107744	2	AC013525	Homo sapi
c	3	18.8	81.7	139563	9	HSJ666F24	Human DNA
	4	18.8	81.7	148184	2	AC017777	Drosophill
	5	18.8	81.7	155047	2	AL390120	Homo sapi
c	6	18.8	81.7	177908	2	AC073977	Homo sapi
	7	18.8	81.7	195755	3	AC010009	Drosophill
	8	18.8	81.7	212739	2	AC018630	Homo sapi
c	9	18.8	81.7	220761	2	AL670227	Mus muscu
	10	18.8	81.7	295566	3	AE003597	Drosophill
	11	18.4	80.0	1305	6	AX122257	Sequence
c	12	18.4	80.0	112618	9	AL390862	Human DNA
	13	18.4	80.0	171943	9	AC012055	Homo sapi
	14	18.4	80.0	179667	2	AC106895	Homo sapi
c	15	18.4	80.0	349880	6	AX127149	Sequence
	16	18.2	79.1	1350	9	HSRNRAGA	H.sapiens m
	17	18.2	79.1	1582	9	BC006433	Homo sapi
	18	18.2	79.1	1607	9	BC009990	Homo sapi
	19	18.2	79.1	1610	9	HSU41654	Human adeno
	20	18.2	79.1	2418	5	CS023186	Crotalus sc
	21	18.2	79.1	2620	5	VAMM12	V.ammodytes
	22	18.2	79.1	2769	5	U01027	Crotalus sc
	23	18.2	79.1	3056	5	VAAAMOL	V.ammodytes
	24	18.2	79.1	3224	5	U01026	Crotalus sc
	25	18.2	79.1	3571	5	TRUGTGPLAB	Trimeresuru
c	26	18.2	79.1	6435	14	PVXX3	Potato virus
	27	18.2	79.1	22204	1	AE008842	Salmonell
c	28	18.2	79.1	57714	2	AC107999	Homo sapi
	29	18.2	79.1	70374	2	AC100649	Mus muscu
	30	18.2	79.1	89177	9	AL356000	Human DNA
c	31	18.2	79.1	110000	2	AL390072_2	Continuation (3 of
	32	18.2	79.1	160626	9	AC006374	Homo sapi
	33	18.2	79.1	160705	9	HS61A9	Human DNA
	34	18.2	79.1	185848	2	AC025929	Homo sapi
	35	18.2	79.1	189371	2	AC023125	Homo sapi
c	36	18.2	79.1	191748	2	AL662902	Mus muscu
	37	18.2	79.1	191772	2	AC022825	Homo sapi
	38	18.2	79.1	193581	2	AL670941	Mus muscu
c	39	18.2	79.1	200475	9	AC079271	Homo sapi
	40	18.2	79.1	200618	9	AC009061	Homo sapi
c	41	18.2	79.1	200724	9	AL139382	Human DNA
	42	18.2	79.1	209262	2	AC021091	Homo sapi
c	43	18.2	79.1	210385	2	AC009627	Homo sapi
	44	18.2	79.1	215399	9	AL160278	Human DNA
c	45	18.2	79.1	225401	2	AL671011	Mus muscu

ALIGNMENTS

RESULT 1

AC069180

LOCUS

DEFINITION

AC069180

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC069130 72241 bp DNA linear HTG 21-MAY-2000
Homo sapiens chromosome 1 clone -2026P8 map 1, LOW-PASS SEQUENCE
SAMPLING.
AC069130
HTG: HTGS_PHASE0.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 72241)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mianga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Olivier, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Rhoman, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigillo, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (21-MAY-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
 Web site: <http://www.seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L6854
 Center clone name: 2026_P_8

* NOTE: This record contains 93 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 658: contig of 658 bp in length
 659 758: gap of 100 bp
 759 1448: contig of 690 bp in length
 1449 1548: gap of 100 bp
 1549 2245: contig of 697 bp in length
 2246 2345: gap of 100 bp
 2346 3039: contig of 694 bp in length
 3040 3139: gap of 100 bp
 3140 3798: contig of 659 bp in length
 3799 3898: gap of 100 bp
 3899 4555: contig of 657 bp in length
 4556 4655: gap of 100 bp
 4656 5337: contig of 682 bp in length
 5338 5437: gap of 100 bp
 5438 6106: contig of 669 bp in length
 6107 6206: gap of 100 bp
 6207 6872: contig of 666 bp in length
 6873 6972: gap of 100 bp
 6973 7651: contig of 679 bp in length
 7652 7751: gap of 100 bp
 7752 8429: contig of 678 bp in length
 8430 8529: gap of 100 bp
 8530 9195: contig of 666 bp in length
 9196 9295: gap of 100 bp
 9296 9965: contig of 670 bp in length
 9966 10065: gap of 100 bp
 10066 10738: contig of 673 bp in length
 10739 10838: gap of 100 bp
 10839 11534: contig of 696 bp in length
 11535 11634: gap of 100 bp
 11635 12340: contig of 706 bp in length
 12341 12440: gap of 100 bp
 12441 13111: contig of 671 bp in length
 13112 13211: gap of 100 bp
 13212 13871: contig of 660 bp in length
 13872 13971: gap of 100 bp
 13972 14641: contig of 670 bp in length
 14642 14741: gap of 100 bp
 14742 15422: contig of 681 bp in length
 15423 15522: gap of 100 bp
 15523 16199: contig of 677 bp in length
 16200 16299: gap of 100 bp
 16300 16973: contig of 674 bp in length
 16974 17073: gap of 100 bp
 17074 17754: contig of 681 bp in length
 17755 17854: gap of 100 bp
 17855 18540: contig of 686 bp in length
 18541 18640: gap of 100 bp
 18641 19293: contig of 653 bp in length
 19294 19393: gap of 100 bp
 19394 20094: contig of 701 bp in length
 20095 20194: gap of 100 bp
 20195 20880: contig of 686 bp in length
 20881 20980: gap of 100 bp
 20981 21662: contig of 682 bp in length
 21663 21762: gap of 100 bp
 21763 22418: contig of 656 bp in length
 22419 22518: gap of 100 bp
 22519 23189: contig of 671 bp in length
 23190 23289: gap of 100 bp
 23290 23934: contig of 645 bp in length
 23935 24034: gap of 100 bp
 24035 24713: contig of 679 bp in length
 24714 24813: gap of 100 bp
 24814 25488: contig of 675 bp in length
 25489 25588: gap of 100 bp
 25589 26261: contig of 673 bp in length
 26262 26361: gap of 100 bp
 26362 27036: contig of 665 bp in length
 27037 27166: gap of 100 bp
 27167 27815: contig of 689 bp in length
 27816 27915: gap of 100 bp
 27916 28577: contig of 662 bp in length
 28578 28677: gap of 100 bp
 28678 29374: contig of 697 bp in length
 29375 29474: gap of 100 bp
 29475 30155: contig of 681 bp in length
 30156 30255: gap of 100 bp
 30256 30942: contig of 687 bp in length
 30943 31042: gap of 100 bp
 31043 31716: contig of 674 bp in length
 31717 31816: gap of 100 bp
 31817 32491: contig of 675 bp in length
 32492 32591: gap of 100 bp
 32592 33266: contig of 675 bp in length
 33267 33366: gap of 100 bp
 33367 34031: contig of 665 bp in length
 34032 34131: gap of 100 bp
 34132 34808: contig of 677 bp in length
 34809 34908: gap of 100 bp
 34909 35587: contig of 679 bp in length
 35588 35687: gap of 100 bp
 35688 36379: contig of 692 bp in length
 36380 36479: gap of 100 bp
 36480 37121: contig of 642 bp in length
 37122 37221: gap of 100 bp
 37222 37915: contig of 694 bp in length
 37916 38015: gap of 100 bp
 38016 38720: contig of 705 bp in length
 38721 38820: gap of 100 bp
 38821 39514: contig of 694 bp in length
 39515 39614: gap of 100 bp
 39615 40284: contig of 670 bp in length
 40285 40384: gap of 100 bp
 40385 41052: contig of 668 bp in length

* 23052 23940: contig of 889 bp in length
 * 23941 24040: gap of 100 bp
 * 24041 24941: contig of 901 bp in length
 * 24942 25041: gap of 100 bp
 * 25042 25965: contig of 924 bp in length
 * 25966 26065: gap of 100 bp
 * 26066 26956: contig of 891 bp in length
 * 26957 27056: gap of 100 bp
 * 27057 27966: contig of 910 bp in length
 * 27967 28066: gap of 100 bp
 * 28067 28973: contig of 907 bp in length
 * 28974 29073: gap of 100 bp
 * 29074 29988: contig of 915 bp in length
 * 29989 30088: gap of 100 bp
 * 30089 31011: contig of 923 bp in length
 * 31012 31111: gap of 100 bp
 * 31112 32020: contig of 909 bp in length
 * 32021 32120: gap of 100 bp
 * 32121 33026: contig of 906 bp in length
 * 33027 33126: gap of 100 bp
 * 33127 34032: contig of 906 bp in length
 * 34033 34132: gap of 100 bp
 * 34133 35038: contig of 906 bp in length
 * 35039 35138: gap of 100 bp
 * 35139 36058: contig of 920 bp in length
 * 36059 36158: gap of 100 bp
 * 36159 37057: contig of 899 bp in length
 * 37058 37157: gap of 100 bp
 * 37158 38058: contig of 901 bp in length
 * 38059 38158: gap of 100 bp
 * 38159 39076: contig of 918 bp in length
 * 39077 39176: gap of 100 bp
 * 39177 40089: contig of 913 bp in length
 * 40090 40189: gap of 100 bp
 * 40190 41105: contig of 916 bp in length
 * 41106 41205: gap of 100 bp
 * 41206 42107: contig of 902 bp in length
 * 42108 42207: gap of 100 bp
 * 42208 43095: contig of 888 bp in length
 * 43096 43195: gap of 100 bp
 * 43196 44090: contig of 895 bp in length
 * 44091 44190: gap of 100 bp
 * 44191 45083: contig of 893 bp in length
 * 45084 45183: gap of 100 bp
 * 45184 46091: contig of 908 bp in length
 * 46092 46191: gap of 100 bp
 * 46192 47091: contig of 900 bp in length
 * 47092 47191: gap of 100 bp
 * 47192 48109: contig of 918 bp in length
 * 48110 48209: gap of 100 bp
 * 48210 49095: contig of 886 bp in length
 * 49096 49195: gap of 100 bp
 * 49196 50097: contig of 902 bp in length
 * 50098 50197: gap of 100 bp
 * 50198 51097: contig of 900 bp in length
 * 51098 51197: gap of 100 bp
 * 51198 52137: contig of 940 bp in length
 * 52138 52237: gap of 100 bp
 * 52238 53142: contig of 905 bp in length
 * 53143 53242: gap of 100 bp
 * 53243 54135: contig of 893 bp in length
 * 54136 54235: gap of 100 bp
 * 54236 55135: contig of 900 bp in length
 * 55136 55235: gap of 100 bp
 * 55236 56107: contig of 872 bp in length
 * 56108 56207: gap of 100 bp
 * 56208 57100: contig of 893 bp in length
 * 57101 57200: gap of 100 bp
 * 57201 58109: contig of 909 bp in length
 * 58110 58209: gap of 100 bp
 * 58210 59113: contig of 904 bp in length
 * 59114 59213: gap of 100 bp
 * 59214 60122: contig of 909 bp in length

* 60123 60222: gap of 100 bp
 * 60223 61095: contig of 873 bp in length
 * 61096 61195: gap of 100 bp
 * 61196 62104: contig of 909 bp in length
 * 62105 62204: gap of 100 bp
 * 62205 63123: contig of 919 bp in length
 * 63124 63223: gap of 100 bp
 * 63224 64141: contig of 918 bp in length
 * 64142 64241: gap of 100 bp
 * 64242 65137: contig of 896 bp in length
 * 65138 65237: gap of 100 bp
 * 65238 66154: contig of 917 bp in length
 * 66155 66254: gap of 100 bp
 * 66255 67090: contig of 836 bp in length
 * 67091 67190: gap of 100 bp
 * 67191 68081: contig of 891 bp in length
 * 68082 68181: gap of 100 bp
 * 68182 69095: contig of 914 bp in length
 * 69096 69195: gap of 100 bp
 * 69196 70073: contig of 878 bp in length
 * 70074 70173: gap of 100 bp
 * 70174 71071: contig of 898 bp in length
 * 71072 71171: gap of 100 bp
 * 71172 72065: contig of 894 bp in length
 * 72066 72165: gap of 100 bp
 * 72166 73083: contig of 918 bp in length

Query Match 81.7%; Score 18.8; DB 2; Length 107744;
 Best Local Similarity 90.9%; Pred. No. 2.6e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 aattcgttcgttcgttttca 23
 | ||||| ||||| ||||| |||||
 Db 39895 ACTTCGCTTGTGCTTTTCA 39916

RESULT 3

HSJ566F24/c
 LOCUS HSJ566F24 139563 bp DNA linear PRI 09-FEB-2002
 DEFINITION Human DNA sequence from clone RP4-666F24 on chromosome lp13.1-13.3,
 complete sequence.
 ACCESSION AL109660.14 GI:18650725
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 139563)
 AUTHORS Frankland, J.
 TITLE Direct Submission
 JOURNAL Submitted (09-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
 On Feb 11, 2002 this sequence version replaced gi:15594038.
 COMMENT During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em:, EMBL; Sw:,
 SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr1>
 RP4-666F24 is from the library RPCI-4 constructed by the group of Pliet de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2

This sequence is the entire insert of clone RP4-666F24 The true left end of clone RP5-977F20 is at 96898 in this sequence. The true right end of clone RP11-109G4 is at 10338 in this sequence.

FEATURES

source

1..139563
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="p13.1-13.3"
 /clone="RP4-666F24"
 /clone_lib="RPCI-4"

misc_feature

87666..87877
 /note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."

BASE COUNT

ORIGIN 40054 a 29037 c 29413 g 41059 t

Query Match

Best Local Similarity 81.7%; Score 18.8; DB 9; Length 139563;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 aattctgcttcgtgttttc 23

| | | | | | | | | | | | | | | | | | | | | |

Db 89983 ACTTCTGCTTGTCTGTTTCA 89962

RESULT 4

AC017777

LOCUS

DEFINITION Drosophila melanogaster, 148184 bp DNA linear HTG 09-DEC-1999 pieces.

AC017777

VERSION AC017777.1 GI:6553413

KEYWORDS HTG: HTGS_PHASE2.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 148184)

Adams,M. and Venter,J.C.

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

This sequence was identified as CDM:10212057 by the submitter.

For more information on this record e-mail to fly@celera.com.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES

source

1..148184

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

BASE COUNT 40862 a 32955 c 32504 g 41863 t

ORIGIN

Query Match

Best Local Similarity 81.7%; Score 18.8; DB 2; Length 148184;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gaattctgcttcgtgttttc 22

| | | | | | | | | | | | | | | | | | | | | |

Db 28035 GAATTGCTGCTTCTGCTTTTC 28056

RESULT 5

AL390120

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (sites)

McLay,K.

Direct Submission

Submitted (20-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Sep 13, 2000 this sequence version replaced gi:10086157.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA521N12

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator ET-amersham; 31% of reads Chemistry;

Dye-terminator Big Dye; 68% of reads

Consensus quality: 140486 bases at least Q40

Consensus quality: 146838 bases at least Q30

Consensus quality: 150087 bases at least Q20

Insert size: 152747; sum-of-contigs

Insert size: 171843; 7.2% error; agarose-fp

Quality coverage: 3.12x in Q20 bases; sum-of-contigs Quality

coverage: 2.96x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 24 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 18738: contig of 18738 bp in length

* 18739 18838: gap of 100 bp

* 18839 22388: contig of 3550 bp in length

* 22389 22488: gap of 100 bp

* 22489 24572: contig of 2084 bp in length

* 24573 24672: gap of 100 bp

* 24673 28174: contig of 3502 bp in length

* 28175 28274: gap of 100 bp

* 28275 30525: contig of 2251 bp in length

* 30526 30625: gap of 100 bp

* 30626 50010: contig of 19385 bp in length

* 50011 50110: gap of 100 bp

* 50111 52255: contig of 2145 bp in length

* 52256 52355: gap of 100 bp

* 52356 58621: contig of 6266 bp in length

* 58622 58721: gap of 100 bp

* 58722 65489: contig of 6768 bp in length

* 65490 65589: gap of 100 bp

* 65590 73034: contig of 7445 bp in length

* 73035 73134: gap of 100 bp

* 73135 76041: contig of 2907 bp in length

* 76042 76141: gap of 100 bp

* 76142 76141: gap of 100 bp

* 76142 76141: gap of 100 bp

* 76142 76141: gap of 100 bp

* 76142 76141: gap of 100 bp

* 76142 76141: gap of 100 bp

* 76142 76141: gap of 100 bp

* 76142 76141: gap of 100 bp

* 76142 76141: gap of 100 bp

* 76142 76141: gap of 100 bp

* 76142 76141: gap of 100 bp

```

source
misc_feature
1..18738
/note="assembly_fragment:00869
fragment_chain:1
clone_end:SP6
vector_side:left"
18839..22388
/note="assembly_fragment:00738
fragment_chain:1"
22489..24572
/note="assembly_fragment:01040
fragment_chain:1"
24673..28174
/note="assembly_fragment:01064
fragment_chain:2"
28275..30525
/note="assembly_fragment:01062
fragment_chain:2"
30626..50010
/note="assembly_fragment:00964
fragment_chain:2"
50111..52255
/note="assembly_fragment:00681
fragment_chain:2"
52356..58621
/note="assembly_fragment:00865
fragment_chain:2"
58722..65489
/note="assembly_fragment:00120
fragment_chain:3"
65590..73034
/note="assembly_fragment:00668
fragment_chain:3"
73135..76041
/note="assembly_fragment:00777
fragment_chain:4"
76142..87642
/note="assembly_fragment:01128
fragment_chain:4"
87743..94188
/note="assembly_fragment:00996
fragment_chain:5"

```

Sequencing vector: M13; 32%
Chemistry: Dye-primer ET; 32% of reads
Assembly: Dye-terminator Big Dye; 67% of reads
Consensus quality: 172974 bases at least Q40
Consensus quality: 174626 bases at least Q30
Insert size: 160000; agarose-fp
Insert size: 177308; sum-of-contigs
Quality coverage: 12.66 in Q20 bases; agarose-fp
Quality coverage: 11.47 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 1146: contig of 1146 bp in length
* 1147 1246: gap of unknown length
* 1247 2420: contig of 1174 bp in length
* 2421 2520: gap of unknown length
* 2521 4176: contig of 1656 bp in length
* 4177 4276: gap of unknown length
* 4277 5991: contig of 1715 bp in length
* 5992 6091: gap of unknown length
* 6092 9107: contig of 3016 bp in length
* 9108 9207: gap of unknown length
* 9208 176414: contig of 167207 bp in length
* 176415 176514: gap of unknown length
* 176515 177908: contig of 1394 bp in length.

FEATURES

source
1. 177908
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-253H16"
1. 1146
/note="assembly_name:Contig21"
1247. 2420
/note="assembly_name:Contig22"
2521. 4176
/note="assembly_name:Contig28"
4277. 5991
/note="assembly_name:Contig29"
6092. 9107
/note="assembly_name:Contig33"
9208. 176414
/note="assembly_name:Contig34"
clone_end:SP6
vector_side:left
176515..177908
/note="assembly_name:Contig6"

BASE COUNT 54431 a 35157 c 34279 g 53411 t 630 others
ORIGIN

Query Match 81.7%; Score 18.8; DB 2; Length 177908;
Best Local Similarity 90.9%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaattcctgttcgtgtttc 22

||||||| ||| |||||

Db 65058 GAATTCCTCTTTCGCTTTC 65037

RESULT 7

AC010009

LOCUS

DEFINITION Drosophila melanogaster 3L BAC RPC198-23D16 (Roswell Park Cancer

195755 bp DNA linear INV 03-JAN-2002

AC010009

LOCUS

DEFINITION

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Institute Drosophila BAC library) complete sequence.

AC010009

AC010009.5 GI:16418028

HTG.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 195755)

Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,

Gocayne,J.D., Tabor,P., Williamson,A., Homsí,F.H.,

Dugan-Rocha,S.D., Sodergren,E.S., Worley,K.M., Amaratunga,P.G., Brannon,R.C.,

Ayala,F., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhaq,C.,

Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,

Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,

Draper,H., Emery-Cohen,A., Ferriera,S., Garg,N.D.S., Houck,J.,

Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,

Liu,W., Mattel,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,

Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,

Scheeler,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,

Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,

Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.

Direct Submission

Unpublished

2 (bases 1 to 195755)

Worley,K.C.

Direct Submission

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

Submitted (25-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

4 (bases 1 to 195755)

AUTHORS

Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsegue,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE

Submitted (27-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

5 (bases 1 to 195755)

AUTHORS

Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsegue,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Moser,M., Mitchell,T., Mohabbat,K.,

TITLE

Submitted (29-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

6 (bases 1 to 195755)

AUTHORS

Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsegue,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Moser,M., Mitchell,T., Mohabbat,K., Morgan,A., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

Query Match 81.7%; Score 18.8; DB 3; Length 195755;
Best Local Similarity 90.9%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaattcctgcgttcgtctgttttc 22

||||| ||||||| |||||||

Db 37603 GAATTGCTGCTTCTCTGTTTC 37624

RESULT

8

AC018630

LOCUS

AC018630 212739 bp DNA linear HTG 13-OCT-2001
DEFINITION Homo sapiens chromosome 12 clone RP11-673D15, WORKING DRAFT

SEQUENCE, 2 unordered pieces.

AC018630

ACCESSION

AC018630 33 GI:16041329

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.

SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS

Mammalia: Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 212739)
Muzly, D.M., Adams, C.C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Anaratunga, H.C., Are, J.R., Banks, T., Barbaria, J.,
Benton, J., Blinige, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhey, C.,
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulsegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
Ogih, M., Okuwon, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Picketts, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojuben, I., Rolfe, M.,
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Taylor, P., Tamerisa, A., Thomas, N.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Groves, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 212739)
Worley, K.C.
Direct Submission
Submitted (15-DEC-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 11, 2001 this sequence version replaced gi:14787084.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMG
Center clone name: RP11-673D15
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 7% of reads
Chemistry: Dye-terminator Big Dye; 87% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 219183 bases at least Q40
Consensus quality: 221416 bases at least Q30
Consensus quality: 222879 bases at least Q20
Estimated insert size: 215919; sum-of-contigs estimation
Quality coverage: 8.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 151372: contig of 151372 bp in length
* 151373 151472: gap of unknown length
* 151473 212739: contig of 61267 bp in length.

FEATURES
Source

Location/Qualifiers
1..212739
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-673D15"
BASE COUNT 70140 a 40132 c 36532 g 65833 t 102 others
ORIGIN

Query Match 81.7%; Score 18.8; DB 2; Length 212739;
Best Local Similarity 90.9%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 gaattcgtgcttcgtgttttc 22
|||||
Db 96992 GAATTCGTCTCCATGTTTC 97013

RESULT 9

AL670227 220761 bp DNA linear HTG 30-JAN-2002
LOCUS Mus musculus chromosome 4 clone RP23-317N1, *** SEQUENCING IN
DEFINITION PROGRESS ***, in unordered pieces.

ACCESSION AL670227

VERSION AL670227.2 GI:18477084

KEYWORDS HTG: HTGS_PHASE1.

SOURCE house mouse.

ORGANISM Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (sites)

Plumb, B

Direct Submission

Submitted (23-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Feb 1, 2002 this sequence version replaced gi:18307364.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BM317N1

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 219697 bases at least Q40

Consensus quality: 219887 bases at least Q30

Consensus quality: 220081 bases at least Q20

Insert size: 220261; sum-of-contigs

Insert size: 213159; 11.3% error; agarose-fp

Quality coverage: 13.01x in Q20 bases; sum-of-contigs Quality

coverage: 13.51x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

Location/Qualifiers

1..220761

/organism="Mus musculus"

FEATURES
source

```

/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-317N1"
/clone_1lb="RPC1-23"
1. .129995
/note="assembly fragment:03451
fragment_chain:1"
130096. .133380
/note="assembly fragment:02708
fragment_chain:1"
133481. .150170
/note="assembly fragment:00061
fragment_chain:1"
150271. .156539
/note="assembly fragment:01326
fragment_chain:1"
156640. .208913
/note="assembly fragment:02733
fragment_chain:1"
209014. .220761
/note="assembly fragment:05258
fragment_chain:1"
clone_end:SP6
vector_side:right"
BASE COUNT 57530 a 54793 c 52961 g 54975 t 502 others
ORIGIN

Query Match 81.7% Score 18.8; DB 2; Length 220761;
Best Local Similarity 90.9%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gaattcctgcctcgtctgttc 22
Db 219218 GAATTCCTGCTCTCTGCTTC 219239

RESULT 10
AE003597/c
LOCUS
DEFINITION
Drosophila melanogaster genomic scaffold 142000013386036 section 7
of 9, complete sequence.
ACCESSION AE003597 AE002647
VERSION AE003597.1 GI:7296525
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 295566)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,J.H., Blazer,R.G., Champ,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evansell,C.C., Ferraz,C., Ferriera,S.,
Fleischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Pan,S.,
Ruskern,D.R., Pacle,J.M., Palazolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirska,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 295566)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
FEATURES
source
location/Qualifiers
1. .295566
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3L"
join(<27808. .27909,28034. .28318,28822. .>29085)
/gene="CG12673"
/product="CT35415"
/db_xref="FLYBASE:FBgn0012673"
/db_xref="FLYBASE:FBgn0037153"
/evidence=not_experimental
<27808. .>29085
/gene="CG12673"
/map="79D4-79E1"
/db_xref="FLYBASE:FBan0012673"
/db_xref="FLYBASE:FBgn0037153"
/evidence=not_experimental
join(27808. .27909,28034. .28318,28822. .29085)
/gene="CG12673"
/note="CG12673 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012673"
/db_xref="FLYBASE:FBgn0037153"
/evidence=not_experimental
/protein_id="AAF51810.1"
/db_xref="GI:7296526"
/translation="MGREGRRAFGPRGAFPAQCIRVHVNSDIIANDTMRLLYMY
HAQDPHGVSVRPGTLPDPAFRPRYPMVLMQRAQLPMPSTHDERVRLVLEINDEVE
LPAGDPLFCMKFLEIDINKKHLIRYEPIYDSSSVHYLQHTLHCOCGAHLEEE
MARQGRPCIGARSIPACNAIVASWGRSEVSISIRNONLPYPPQCNSCPQT"
join(<35333. .35500,35594. .35756,35909. .>35937)
/gene="CG7495"
/product="CT23023"
/db_xref="FLYBASE:FBan0007495"
/db_xref="FLYBASE:FBgn0037154"
<35333. .>35937
/gene="CG7495"
/map="79E1-79E1"
/db_xref="FLYBASE:FBan0007495"
/db_xref="FLYBASE:FBgn0037154"
join(35333. .35500,35594. .35756,35909. .35937)
/gene="CG7495"
/note="CG7495 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0007495"
/db_xref="FLYBASE:FBgn0037154"
/protein_id="AAF51811.1"
/db_xref="GI:7296527"
/translation="MDPNWRHIIPPQKRVVSEQCIEDCTGYAPPOGQINIFAVMMR"

```



```

SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 112618)
JOURNAL     Tracey, A.
COMMENT     Direct Submission
            Submitted (11-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequests@sanger.ac.uk
            On Oct 12, 2001 this sequence version replaced gi:14787516.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/projects/C_elegans/wormpep
            This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-30G13 is from the library RPi-11.1 constructed by the group
            of Pictet de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACE3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-30G13. It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-103A2 is at 110619 in this
            sequence. The true right end of clone RP11-94M14 is at 2000 in this
            sequence.

FEATURES             Location/Qualifiers
     source           1..112618
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-30G13"
                     /clone_lib="RP11-30G13"
                     /clone_lib="RP11-30G13"
                     /note="Single clone region. Assembly confirmed by
                     restriction digest data."
     misc_feature     9227..9232
                     /note="Single clone region. Assembly confirmed by
                     restriction digest data."
     BASE COUNT      36195 a 21831 c 20733 g 33859 t
     ORIGIN
     Query Match      80.0%; Score 18.4; DB 9; Length 112618;
     Best Local Similarity 95.0%; Pred. NO. 4e+02;
     Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aattctgcttcgtgtttt 21
|||||
DB 26288 AATTCTGCTTCTCTCTGTTT 26269

RESULT 13
AC012055/c
LOCUS      AC012055      171943 bp      DNA      linear      PRI 30-JUN-2000
DEFINITION Homo sapiens chromosome 4, clone RP11-248N22, complete sequence.
ACCESSION AC012055
VERSION   AC012055.9  GI:8671944
KEYWORDS  HTG.

```

```

SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 171943)
JOURNAL     Birren, B., Linton, L., Nusbaum, C. and Lander, E.
COMMENT     Homo sapiens chromosome 4, clone RP11-248N22
            Unpublished
            2 (bases 1 to 171943)
            Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
            Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B.,
            Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
            Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
            Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
            Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
            Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
            Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
            McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
            Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
            Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
            Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
            Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
            Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
            Direct Submission
            Submitted (19-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            3 (bases 1 to 171943)
            Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
            Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
            Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
            Campoliano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
            Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
            Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
            Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
            Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
            Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
            Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
            Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
            McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
            Meldrim, J., Menues, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
            Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
            O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
            Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
            Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
            Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
            Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
            Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
            Young, G., Zainoun, J., Zimmer, A. and Zody, M.
            Direct Submission
            Submitted (30-JUN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Jun 23, 2000 this sequence version replaced gi:8569149.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L1734
            Center clone name: 248_N_22
            ----- Location/Qualifiers
            1..171943
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="4"
            /map="4"
            /clone="RP11-248N22"
            /clone_lib="RP11-248N22"
            complement(2..780)
            repeat_region

```



```
repeat_region /rpt_family="L1PBI"
repeat_region /rpt_family="LTR40b"
repeat_region /rpt_family="L2"
repeat_region complement(2269. .2496)
repeat_region /rpt_family="L2"
repeat_region 2578. .2710
repeat_region /rpt_family="L2"
repeat_region complement(2724. .2800)
repeat_region /rpt_family="L2"
repeat_region 4277. .4454
repeat_region /rpt_family="MIR"
repeat_region 4456. .4703
repeat_region /rpt_family="LMC5"
repeat_region complement(5952. .6254)
repeat_region /rpt_family="AluSq"
repeat_region 6257. .6279
repeat_region /rpt_family="AT_rich"
repeat_region complement(6280. .6568)
repeat_region /rpt_family="AluSc"
repeat_region 6569. .6585
repeat_region /rpt_family="AT_rich"
repeat_region complement(7177. .7226)
repeat_region /rpt_family="NSTA"
repeat_region 7617. .7665
repeat_region /rpt_family="AT_rich"
repeat_region complement(8417. .8633)
repeat_region /rpt_family="AluSq"
repeat_region 10488. .10560
repeat_region /rpt_family="MIR"
repeat_region complement(11030. .11134)
repeat_region /rpt_family="MLT1B"
repeat_region 11135. .11307
repeat_region /rpt_family="MER41B"
repeat_region 11308. .11536
repeat_region /rpt_family="L1PA7"
repeat_region 11537. .11912
repeat_region /rpt_family="MER41B"
repeat_region complement(11925. .12155)
repeat_region /rpt_family="AluSq"
repeat_region 12179. .12302
repeat_region /rpt_family="MER41B"
repeat_region complement(12312. .12556)
repeat_region /rpt_family="MLT1B"
repeat_region complement(12773. .13068)
repeat_region /rpt_family="AluSx"
repeat_region 13453. .13530
repeat_region /rpt_family="(TATAA)n"
repeat_region 16473. .16498
repeat_region /rpt_family="(TG)n"
repeat_region complement(17063. .17391)
repeat_region /rpt_family="AluSq"
repeat_region complement(17608. .17760)
repeat_region /rpt_family="L1M4"
repeat_region 17910. .17968
repeat_region /rpt_family="MIR"
repeat_region 18124. .18417
repeat_region /rpt_family="AluSx"
repeat_region 18634. .18677
repeat_region /rpt_family="AT_rich"
repeat_region 19011. .19032
repeat_region /rpt_family="AT_rich"
repeat_region 19283. .19412
repeat_region /rpt_family="MIR"
repeat_region complement(21374. .21657)
repeat_region /rpt_family="AluSg"
repeat_region 23426. .23553
repeat_region /rpt_family="MIR"
repeat_region 23843. .23864
repeat_region /rpt_family="AT_rich"
repeat_region complement(24069. .24159)
repeat_region /rpt_family="MLT1K"
```

```
repeat_region complement(24164. .24289)
repeat_region /rpt_family="AluJb"
repeat_region 25593. .25631
repeat_region /rpt_family="(TTTA)n"
repeat_region complement(25849. .26021)
repeat_region /rpt_family="MIR"
repeat_region 26807. .27161
repeat_region /rpt_family="THE1B"
repeat_region 28665. .29002
repeat_region /rpt_family="MLT1A1"
repeat_region 31978. .32056
repeat_region /rpt_family="L1MC4"
repeat_region 32109. .32328
repeat_region /rpt_family="L1MC4"
repeat_region 33299. .33452
repeat_region /rpt_family="(TATAA)n"
repeat_region 33513. .33577
repeat_region /rpt_family="AT_rich"
repeat_region 34264. .34297
repeat_region /rpt_family="(TTTTTG)n"
repeat_region 35484. .35507
repeat_region /rpt_family="AT_rich"
repeat_region 36140. .36186
repeat_region /rpt_family="AT_rich"
repeat_region 36284. .36312
repeat_region /rpt_family="AT_rich"
repeat_region 37765. .37791
repeat_region /rpt_family="AT_rich"
repeat_region 37856. .37878
repeat_region /rpt_family="AT_rich"
repeat_region 38007. .38035
repeat_region /rpt_family="AT_rich"
repeat_region complement(40978. .41140)
repeat_region /rpt_family="MER103"
repeat_region 41297. .41330
repeat_region /rpt_family="(TAGA)n"
repeat_region complement(41789. .41848)
repeat_region /rpt_family="L2"
repeat_region complement(41803. .41938)
repeat_region /rpt_family="MIR"
repeat_region 43710. .43851
repeat_region /rpt_family="MIR"
repeat_region 43958. .44767
repeat_region /rpt_family="L1ME3"
repeat_region 45733. .45761
repeat_region /rpt_family="AT_rich"
repeat_region complement(47019. .47316)
repeat_region /rpt_family="AluY"
repeat_region 48229. .48276
```

Query Match 80.0%; Score 18.4; DB 9; Length 171943;

Best Local Similarity 95.0%; Pred. NO. 3.8e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aattctgcttcgctgtttt 21

||||| |||||||

Db 170336 AATTCTACTTCGCTGTTT 170317

RESULT 14

AC106895/c

LOCUS AC106895 Homo sapiens chromosome 4 clone RP11-161D15, WORKING DRAFT

DEFINITION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

ACCESSION AC106895

VERSION AC106895.3 GI:18308813

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 179667)

AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179667)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (13-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Jan 24, 2002 this sequence version replaced gi:18141525.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: H_NH0161D15
----- Summary Statistics -----
Sequencing vector: M13: 0%
Chemistry: Dye-primer ET; 100% of reads
Assembly: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 178624 bases at least Q40
Consensus quality: 178982 bases at least Q30
Consensus quality: 179266 bases at least Q20
Insert size: 173000; agarose-fp
Insert size: 179567; sum-of-contigs
Quality coverage: 9.03 in Q20 bases; agarose-fp
Quality coverage: 8.15 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 65346: contig of 65346 bp in length
* 65347 65446: gap of unknown length
* 65447 179667: contig of 114221 bp in length.

FEATURES

source
1..179667
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-161D15"

misc_feature
1..65346
/note="assembly_name:Contig98"
misc_feature
65447..179667
/note="assembly_name:Contig99"
clone_end:SP6
vector_side:right

BASE COUNT 57454 a 33427 c 31929 g 56757 t 100 others
ORIGIN

Query Match 80.0%; Score 18.4; DB 2; Length 179667;
Best Local Similarity 95.0%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 aattcctgcttcgctgtttt 21
||||| |||||||
Db 44493 AATTCTACTTCGCTGTTT 44474

RESULT 15
AX127149/c 349980 bp DNA linear PAT 11-MAY-2001
LOCUS AX127149
DEFINITION Sequence 7065 from Patent EP1108790.
ACCESSION AX127149 AX114121

VERSION
KEYWORDS
SOURCE
ORGANISM

AX127149.1 GI:14041137
Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
Corynebacterium.

REFERENCE

1 (bases 1 to 349980)
AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
TITLE Novel polynucleotides

JOURNAL Patent: EP 1108790-A 7065 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)

FEATURES

source
1..349980
Location/Qualifiers
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
/note="Seq 1 to long (3.309.400) split in 11, seq 7065
1.800.001 2.149.980"

BASE COUNT 86896 a 98023 c 80939 g 84122 t
ORIGIN

Query Match 80.0%; Score 18.4; DB 6; Length 349980;
Best Local Similarity 95.0%; Pred. No. 3.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY

2 aattcctgcttcgctgtttt 21
||||| |||||||
Db 291447 AATTCTACTTCGCTGTTT 291428

Search completed: July 25, 2002, 05:39:24
Job time: 6717 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 05:44:43 ; Search time 378.07 Seconds
(without alignments)
104.449 Million cell updates/sec

Title: US-09-235-416-4
Perfect score: 23
Sequence: 1 gaattctgcttcgtgtttca 23

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	100.0	23	AAH87658	Kinesin motor prot
C 2	23	100.0	2352	AAH87656	Thermomyces lanugi
C 3	18.8	81.7	466	AA216520	Human gene express
C 4	18.4	80.0	1305	AAH67138	C glutamicum codin
C 5	18.4	80.0	349980	AAH68530	C glutamicum codin
C 6	18.2	79.1	705	AAH56002	Salmonella typhi D
C 7	17.8	77.4	2408	AB199592	Mouse ischaemic co
C 8	17.4	75.7	2288	AAD08438	Human secreted pro
C 9	17.4	75.7	8734	ABA09012	Human TRAP240 homo

ALIGNMENTS

RESULT 1

AAH87658

ID AAX87658 standard; DNA; 23 BP.

XX AC AAX87658;

XX DT 26-OCT-1999 (first entry)

XX DE Kinesin motor protein TL-gamma gene PCR primer.

XX KW TL-gamma; kinesin; motor protein; microtubule; unc-104; infection;

XX KW neurodegenerative disease; Alzheimer's disease;

XX KW Parkinson's disease; Huntington's disease;

XX KW anyotrophic lateral sclerosis; PCR; primer; ss.

XX OS Synthetic.

XX OS Thermomyces lanuginosus.

XX PN WO9937659-A1.

XX PD 29-JUL-1999.

XX PF 22-JAN-1999; 99WO-US01355.

XX PR 23-JAN-1998; 98US-0072361.

XX PA (REGC) UNIV CALIFORNIA.

XX XX Goldstein LSB, Sakowicz R;

XX XX WPI; 1999-493950/41.

C 10	17.4	75.7	8734	22	AAI59594	Human polynucleoti
C 11	17.4	75.7	8761	22	AAI57808	Human polynucleoti
C 12	17.4	75.7	8785	23	AAH85661	DNA encoding novel
C 13	17.2	74.8	4016	22	AAK77376	Human immune/haema
C 14	17.2	74.8	4762	22	AAK77377	Human immune/haema
C 15	16.8	73.0	360	19	AAH14159	H. pylori GHPO 491
C 16	16.8	73.0	4078	22	AAH16633	Human cDNA sequenc
C 17	16.8	73.0	4715	22	ABAI6034	Human nervous syst
C 18	16.8	73.0	4715	22	ABAI6035	Human nervous syst
C 19	16.8	73.0	127197	22	ABAI61370	Soybean 515002 reg
C 20	16.6	72.2	212	21	AAH03476	Human secreted pro
C 21	16.6	72.2	216	20	AAH03476	Human secreted pro
C 22	16.6	72.2	393	20	AAH20106	Enterococcus faeca
C 23	16.6	72.2	538	22	AAH40008	Genomic sequence #
C 24	16.6	72.2	538	22	AAH40008	Human digestive sy
C 25	16.6	72.2	1965	23	AAH66737	DNA encoding novel
C 26	16.6	72.2	4758	18	AAH96561	Rat neurodapl 1 gen
C 27	16.6	72.2	4778	18	AAH96562	Human neurodapl 1 g
C 28	16.6	72.2	20561	20	AAH13329	Enterococcus faeca
C 29	16.6	72.2	58708	22	AAH64739	Human immune/haema
C 30	16.6	72.2	72604	20	AAH10752	Genomic sequence o
C 31	16.6	72.2	119950	20	AAH90201	Human yes1 gene.
C 32	16.6	72.2	236303	22	AAH11614	Human genomic DNA
C 33	16.4	71.3	31	20	AAH39026	Human genomic DNA
C 34	16.2	70.4	196	21	AAH43132	Human secreted exp
C 35	16.2	70.4	198	21	AAH08506	Human secreted pro
C 36	16.2	70.4	406	21	AAH21446	Human secreted pro
C 37	16.2	70.4	412	21	AAH43726	Mouse secreted exp
C 38	16.2	70.4	485	22	AAH24961	Human ovarian PCR-
C 39	16.2	70.4	485	22	AAH83606	Human ovarian tumo
C 40	16.2	70.4	504	20	AAH88728	EST clone HJ181.
C 41	16.2	70.4	1667	21	AAH62532	Human secreted pro
C 42	16.2	70.4	1667	22	AAH22706	Human secreted pro
C 43	16.2	70.4	2345	19	AAH71209	DNA encoding ester
C 44	16.2	70.4	2405	19	AAH71215	DNA encoding ester
C 45	16.2	70.4	3155	24	AAH94847	Human DNA sequence

XX The present invention describes a library of human polynucleotides
CC comprising the sequences given in AA212532 to AA217779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.

XX SQ Sequence 466 BP; 185 A; 65 C; 97 G; 116 T; 3 other;

Query Match 81.7%; Score 18.8; DB 20; Length 466;
Best Local Similarity 90.9%; Pred. No. 41;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gaattctgctgctgctgtttc 22
|||||
Db 434 GAATTCCTGCTCCATGTTTC 413

RESULT 4
AAH67138
ID AAH67138 standard; DNA; 1305 BP.
AC AAH67138;
DT 26-SEP-2001 (first entry)
DE C glutamicum coding sequence fragment SEQ ID NO: 2173.
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX Corynebacterium glutamicum.
OS
PN EPI108790-A2.
XX 20-JUN-2001.
PD
PF 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.
DR P-PSDB; AAG91919.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene

PS Claim 8; SEQ ID NO: 2173; 246pp + Sequence Listing; English.
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

XX SQ Sequence 1305 BP; 232 A; 298 C; 389 G; 386 T; 0 other;

Query Match 80.0%; Score 18.4; DB 22; Length 1305;
Best Local Similarity 95.0%; Pred. No. 67;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aattctgctgctgctgtttt 21
|||||
Db 609 aattactgctgctgctgtttt 628

RESULT 5
AAH68530/c
ID AAH68530 standard; DNA; 349980 BP.
XX
AC AAH68530;
DT 26-SEP-2001 (first entry)
DE C glutamicum coding sequence fragment SEQ ID NO: 7065.
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX Corynebacterium glutamicum.
OS
PN EPI108790-A2.
XX 20-JUN-2001.
PD
PF 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene

PS Disclosure; SEQ ID NO: 7065; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived

SQ Sequence 2408 BP; 694 A; 501 C; 635 G; 578 T; 0 other;

Query Match 77.4%; Score 17.8; DB 24; Length 2408;
Best Local Similarity 90.5%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 attctgcttcgctgtttca 23
||||| ||||| ||| |||||

Db 211 attctgcttcgctgtttca 231

RESULT 8
AAD08438/c
ID AAD08438 standard; cDNA; 2288 BP.

XX
AC AAD08438;
DT
DT
XX 09-AUG-2001 (first entry)

DE Human secreted protein-encoding gene 5 cDNA clone HLDB47, SEQ ID NO:45.

XX Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammatory; neurological disorder; Alzheimer's disease; food additive;
KW angiotensin disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vulnery; binding partner identification;
KW gene therapy; chromosome 4; ss.

XX Homo sapiens.

OS
FH Key Location/Qualifiers
FT CDS complement (1302..1622)
FT FT /*tag= a
FT FT /product= "Human secreted protein precursor"
FT FT /note= "CDS does not include start codon"
FT FT /partial
FT FT sig_peptide complement (1620..1622)
FT FT /*tag= b
FT FT mat_peptide complement (1305..1619)
FT FT /*tag= c
FT FT /product= "Mature human secreted protein"

XX WO200134643-A1.
XX
XX
XX 17-MAY-2001.
XX
XX 08-NOV-2000; 2000WO-US30629.
XX
XX 12-NOV-1999; 99US-0164825.
PR 03-AUG-2000; 2000US-0222904.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
PI Ruben SM, Komatsoulis GA, Soppet DR, Shi Y;
XX WPI: 2001-374441/39.
DR P-PSDB; AAE04133.
XX
XX Nucleic acids encoding 24 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. Gaucher's disease,
PT Alzheimer's disease, Scimitar syndrome, Creutzfeldt-Jacob disease,
PT diabetes mellitus and multiple sclerosis -
XX
PS Claim 1: Page 441; 532pp; English.

XX
XX AAD08404-AAD08478 represent cDNAs corresponding to 24 human secreted
CC protein genes, and AAE04100-AAE004170 represent the proteins they encode.
CC AAE04172-AAE04197 represent human secreted protein fragments or variants.

CC The secreted proteins and their genes are useful for preventing, treating
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 24 genes.
CC based on the tissues in which they are most highly expressed, and include
CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC angiotensin-related disorders, kidney disorders, gastrointestinal disorders,
CC pregnancy-related disorders, endocrine disorders, and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
CC The present sequence represents a human secreted protein-encoding cDNA of
CC the invention.
XX
SQ Sequence 2288 BP; 664 A; 491 C; 456 G; 672 T; 5 other;

Query Match 75.7%; Score 17.4; DB 22; Length 2288;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gaattctcgtcgtcgtgtt 19
||||| ||||| |||||

Db 480 GAATTCCTGCTCGCTGTT 462

RESULT 9
ABA09012/c
ID ABA09012 standard; cDNA; 8734 BP.

XX ABA09012;
AC ABA09012;
XX
DT 11-JAN-2002 (first entry)
XX
XX Human TRAP240 homologue-encoding cDNA, SEQ ID NO:788.

XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antiulcer; ss.

XX Homo sapiens.
OS
XX
XX WO200157188-A2.
XX
XX 09-AUG-2001.
PD
XX
XX 05-FEB-2001; 2001WO-US03800.
PF
XX
XX 03-FEB-2000; 2000US-0496914.
PR
XX 27-APR-2000; 2000US-0560875.

XX DT 22-OCT-2001 (first entry)
XX DE Human polynucleotide SEQ ID NO 11.
XX AC
XX DT
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX OS Homo sapiens.
XX PN WO200153312-A1.
XX PD
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
XX P-PSDB; AAM38652.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Claim 1; SEQ ID NO 11; 10078pp; English.
XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX SQ Sequence 8761 BP; 2551 A; 1778 C; 1784 G; 2648 T; 0 other;

Query Match 75.7%; Score 17.4; DB 22; Length 8761;
Best Local Similarity 94.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gaattctcgttcgctgtt 19
|||||
DB 3144 GAATTCCTGCTCGCTGTT 3126

RESULT 12

AAS85661/c
ID AAS85661 standard; cDNA; 8785 BP.
XX AC
XX AC AAS85661;
XX DT
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #21465.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX P-PSDB; ABG21474.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX PS Claim 1; SEQ ID NO 21465; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 8785 BP; 2531 A; 1811 C; 1806 G; 2637 T; 0 other;

Query Match 75.7%; Score 17.4; DB 23; Length 8785;
Best Local Similarity 94.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gaattctcgttcgctgtt 19
|||||
DB 3104 GAATTCCTGCTCGCTGTT 3086

RESULT 13
AAK77376/c

ID	AAK77376 standard; DNA; 4016 BP.	PR	14-SEP-2000; 2000US-0232399.
XX		PR	14-SEP-2000; 2000US-0232400.
AC	AAK77376;	PR	14-SEP-2000; 2000US-0232401.
XX		PR	14-SEP-2000; 2000US-0233063.
DT	07-NOV-2001 (first entry)	PR	14-SEP-2000; 2000US-0233064.
XX		PR	14-SEP-2000; 2000US-0233065.
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32188.	PR	21-SEP-2000; 2000US-0234223.
XX		PR	21-SEP-2000; 2000US-0234274.
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;	PR	25-SEP-2000; 2000US-0234997.
KW	cytostatic; gene therapy; vaccine; metastasis; ds.	PR	25-SEP-2000; 2000US-0234998.
OS	Homo sapiens.	PR	26-SEP-2000; 2000US-0235484.
XX		PR	27-SEP-2000; 2000US-0235834.
PN	WO200157182-A2.	PR	27-SEP-2000; 2000US-0235836.
XX		PR	29-SEP-2000; 2000US-0236327.
XX		PR	29-SEP-2000; 2000US-0236367.
PD	09-AUG-2001.	PR	29-SEP-2000; 2000US-0236368.
XX		PR	29-SEP-2000; 2000US-0236369.
XX		PR	29-SEP-2000; 2000US-0236370.
PF	17-JAN-2001; 2001WO-US01354.	PR	02-OCT-2000; 2000US-0236802.
XX		PR	02-OCT-2000; 2000US-0237037.
XX	31-JAN-2000; 2000US-0179065.	PR	02-OCT-2000; 2000US-0237038.
PR	04-FEB-2000; 2000US-0180628.	PR	02-OCT-2000; 2000US-0237039.
PR	24-FEB-2000; 2000US-0184664.	PR	02-OCT-2000; 2000US-0237040.
PR	02-MAR-2000; 2000US-0186350.	PR	02-OCT-2000; 2000US-0239935.
PR	16-MAR-2000; 2000US-0189874.	PR	13-OCT-2000; 2000US-0239935.
PR	17-MAR-2000; 2000US-0190076.	PR	13-OCT-2000; 2000US-0239937.
PR	18-APR-2000; 2000US-0198123.	PR	20-OCT-2000; 2000US-0240960.
PR	19-MAY-2000; 2000US-0205515.	PR	20-OCT-2000; 2000US-0241221.
PR	07-JUN-2000; 2000US-0209467.	PR	20-OCT-2000; 2000US-0241785.
PR	28-JUN-2000; 2000US-0214886.	PR	20-OCT-2000; 2000US-0241786.
PR	30-JUN-2000; 2000US-0215135.	PR	20-OCT-2000; 2000US-0241787.
PR	07-JUL-2000; 2000US-0216647.	PR	20-OCT-2000; 2000US-0241808.
PR	07-JUL-2000; 2000US-0216680.	PR	20-OCT-2000; 2000US-0241809.
PR	11-JUL-2000; 2000US-0217487.	PR	20-OCT-2000; 2000US-0241826.
PR	11-JUL-2000; 2000US-0217496.	PR	01-NOV-2000; 2000US-0244617.
PR	14-JUL-2000; 2000US-0218290.	PR	08-NOV-2000; 2000US-0246474.
PR	26-JUL-2000; 2000US-0220963.	PR	08-NOV-2000; 2000US-0246475.
PR	26-JUL-2000; 2000US-0220964.	PR	08-NOV-2000; 2000US-0246476.
PR	14-AUG-2000; 2000US-0224518.	PR	08-NOV-2000; 2000US-0246477.
PR	14-AUG-2000; 2000US-0224519.	PR	08-NOV-2000; 2000US-0246478.
PR	14-AUG-2000; 2000US-0225213.	PR	08-NOV-2000; 2000US-0246523.
PR	14-AUG-2000; 2000US-0225214.	PR	08-NOV-2000; 2000US-0246524.
PR	14-AUG-2000; 2000US-0225266.	PR	08-NOV-2000; 2000US-0246525.
PR	14-AUG-2000; 2000US-0225267.	PR	08-NOV-2000; 2000US-0246526.
PR	14-AUG-2000; 2000US-0225268.	PR	08-NOV-2000; 2000US-0246527.
PR	14-AUG-2000; 2000US-0225270.	PR	08-NOV-2000; 2000US-0246528.
PR	14-AUG-2000; 2000US-0225447.	PR	08-NOV-2000; 2000US-0246532.
PR	14-AUG-2000; 2000US-0225757.	PR	08-NOV-2000; 2000US-0246609.
PR	14-AUG-2000; 2000US-0225758.	PR	08-NOV-2000; 2000US-0246610.
PR	14-AUG-2000; 2000US-0225759.	PR	08-NOV-2000; 2000US-0246611.
PR	18-AUG-2000; 2000US-0226279.	PR	08-NOV-2000; 2000US-0246613.
PR	22-AUG-2000; 2000US-0226681.	PR	17-NOV-2000; 2000US-0249207.
PR	22-AUG-2000; 2000US-0226686.	PR	17-NOV-2000; 2000US-0249208.
PR	22-AUG-2000; 2000US-0227182.	PR	17-NOV-2000; 2000US-0249209.
PR	23-AUG-2000; 2000US-0227009.	PR	17-NOV-2000; 2000US-0249210.
PR	30-AUG-2000; 2000US-0228924.	PR	17-NOV-2000; 2000US-0249211.
PR	01-SEP-2000; 2000US-0229287.	PR	17-NOV-2000; 2000US-0249212.
PR	01-SEP-2000; 2000US-0229343.	PR	17-NOV-2000; 2000US-0249213.
PR	01-SEP-2000; 2000US-0229344.	PR	17-NOV-2000; 2000US-0249214.
PR	01-SEP-2000; 2000US-0229345.	PR	17-NOV-2000; 2000US-0249215.
PR	05-SEP-2000; 2000US-0229509.	PR	17-NOV-2000; 2000US-0249216.
PR	05-SEP-2000; 2000US-0229513.	PR	17-NOV-2000; 2000US-0249217.
PR	06-SEP-2000; 2000US-0230437.	PR	17-NOV-2000; 2000US-0249218.
PR	06-SEP-2000; 2000US-0230438.	PR	17-NOV-2000; 2000US-0249244.
PR	08-SEP-2000; 2000US-0231242.	PR	17-NOV-2000; 2000US-0249245.
PR	08-SEP-2000; 2000US-0231243.	PR	17-NOV-2000; 2000US-0249264.
PR	08-SEP-2000; 2000US-0231244.	PR	17-NOV-2000; 2000US-0249265.
PR	08-SEP-2000; 2000US-0231413.	PR	17-NOV-2000; 2000US-0249297.
PR	08-SEP-2000; 2000US-0231414.	PR	17-NOV-2000; 2000US-0249299.
PR	08-SEP-2000; 2000US-0232080.	PR	17-NOV-2000; 2000US-0249300.
PR	08-SEP-2000; 2000US-0232081.	PR	01-DEC-2000; 2000US-0250160.
PR	12-SEP-2000; 2000US-0231968.	PR	01-DEC-2000; 2000US-0250391.
PR	14-SEP-2000; 2000US-0232397.	PR	05-DEC-2000; 2000US-0251030.
PR	14-SEP-2000; 2000US-0232398.	PR	05-DEC-2000; 2000US-0251988.

```
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PA Rosen CA, Barash SC, Ruben SM;
XX PI WPI; 2001-483426/52.
XX DR
XX PT Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
XX PT useful for preventing, diagnosing and/or treating cancers and
XX PT metastasis -
XX
XX PS Disclosure; SEQ ID NO 32188; 3071pp + Sequence Listing; English.
XX
XX CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX CC activity, and can be used in gene therapy and vaccine production. (I)
XX CC proteins and polynucleotides may be used in the prevention, diagnosis and
XX CC treatment of diseases associated with inappropriate (I) expression. For
XX CC example, they may be used to treat disorders associated with decreased
XX CC expression by rectifying mutations or deletions in a patient's genome
XX CC that affect the activity of (I) by expressing inactive proteins or to
XX CC supplement the patients own production of (I). Additionally, (I)
XX CC polynucleotides may be used to produce the secreted (I), by inserting
XX CC the nucleic acids into a host cell and culturing the cell to express the
XX CC protein. (I) proteins and polynucleotides may be used to prevent,
XX CC diagnose and treat immune/haematopoietic-related diseases, especially
XX CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX CC to AAK87694 represent human immune/haematopoietic antigen genomic
XX CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX CC represent sequences used in the exemplification of the present invention.
XX
XX SQ Sequence 4016 BP; 1298 A; 763 C; 795 G; 1160 T; 0 other;

Query Match 74.8%; Score 17.2; DB 22; Length 4016;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aattcctgcttcgttttca 23
    ||||| ||||| ||||| |||||
DB 2175 AATTTCGCTTACCTGTTTCA 2154

RESULT 14
AAK77377/c
ID AAK77377 standard; DNA; 4762 BP.
XX AC AAK77377;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32189.
DE
XX Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
PN
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
```

```
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216847.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
```

PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
PR (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -

XX Disclosure; SEQ ID NO 32189; 3071pp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 4762 BP; 1540 A; 898 C; 947 G; 1377 T; 0 other;

Query Match 74.8%; Score 17.2; DB 22; Length 4762;
Best Local Similarity 86.4%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aattctgcttgcgtgtttca 23
||||| ||||| ||||| |||||
Db 2923 AATTTCGCTTACCTGTTTCA 2902

RESULT 15
AAK14159/C
ID AAK14159 standard; DNA; 360 BP.
XX
AC AAK14159;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 491 gene.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 82..297
FT /*tag= a
XX
XX WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
XX 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
XX Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX WPI; 1998-542293/46.
DR P-PSDB; AAW98440.
XX
XX New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases

XX Claim 1; Page 756; 2054pp; English.
PS
XX
CC This sequence represents a polynucleotide of the invention. It was
CC isolated from *Helicobacter pylori* and encodes a *H. pylori* GHPO protein.
CC The polypeptides can be used for preventing or treating *Helicobacter*
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 360 BP; 160 A; 64 C; 71 G; 65 T; 0 other;

Query Match 73.0%; Score 16.8; DB 19; Length 360;
Best Local Similarity 90.0%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 4 ttcttgcttcgctgtttcca 23
DB 169 TTAGTCTTCGCTGTTTCA 150

Search completed: July 25, 2002, 05:44:57
Job time: 6965 sec

THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description	
	Score	Match	Length				
c	1	16.6	72.2	72604	4	US-09-268-992-7	
	2	16.2	70.4	2345	4	US-09-058-260-15	
	3	16.2	70.4	2405	4	US-09-058-260-27	
	4	16.2	70.4	3513	4	US-09-058-260-1	
	5	16.2	70.4	3545	2	US-08-781-802-9	
	6	16.2	70.4	4315	2	US-08-781-802-1	
	7	16.2	70.4	4315	4	US-08-694-078-1	
	8	15.8	68.7	752	3	US-09-008-979A-8	
	9	15.8	68.7	752	4	US-09-460-618-8	
	10	15.6	67.8	431	1	US-08-474-140-15	
	11	15.6	67.8	431	1	US-08-477-630-15	
	12	15.6	67.8	431	1	US-08-472-293-15	
	13	15.6	67.8	431	1	US-08-474-545-15	
	14	15.6	67.8	431	1	US-08-478-341-15	
	15	15.6	67.8	431	3	US-08-996-733-15	
	16	15.6	67.8	889	1	US-08-832-883-52	
	c	17	15.6	67.8	889	2	US-08-832-877-52
	c	18	15.6	67.8	994	3	US-08-961-083-211
	19	15.6	67.8	1033	3	US-08-961-083-191	
	20	15.6	67.8	4464	1	US-08-474-140-8	
	21	15.6	67.8	4464	1	US-08-474-140-9	
	22	15.6	67.8	4464	1	US-08-477-630-8	
	23	15.6	67.8	4464	1	US-08-477-630-9	
	24	15.6	67.8	4464	1	US-08-472-293-8	
	25	15.6	67.8	4464	1	US-08-472-293-9	
	26	15.6	67.8	4464	1	US-08-474-545-8	
	27	15.6	67.8	4464	1	US-08-474-545-9	

RESULT 2
US-09-058-260-15
; Sequence 15, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
; APPLICANT: Allen, Larry
; APPLICANT: Atkins, John
; APPLICANT: Fonsteln, Michael

```
;
; APPLICANT: Vonstein, Veronika
; APPLICANT: Demirjian, David
; APPLICANT: Casadaban, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; FILE REFERENCE: 95-963-H
; CURRENT APPLICATION NUMBER: US/09/058,260B
; CURRENT FILING DATE: 1999-04-10
; EARLIER APPLICATION NUMBER: 60/001,995
; EARLIER FILING DATE: 1996-08-07
; EARLIER APPLICATION NUMBER: 60/009,704
; EARLIER FILING DATE: 1996-01-11
; EARLIER APPLICATION NUMBER: 60/019,580
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: 08/694,078
; EARLIER FILING DATE: 1996-08-08
; EARLIER APPLICATION NUMBER: 08/781,802
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 08/827,810
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 2345
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
; OTHER INFORMATION: gene from bacteria E006
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (196)..(1689)
; US-09-058-260-15

Query Match          70.4%; Score 16.2; DB 4; Length 2345;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattctgcttcgctgtttt 21
    ||||| ||| ||| ||||| |||||
Db 2124 gaatccctccatcgctgtttt 2144

RESULT 3
US-09-058-260-27
; Sequence 27, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
; APPLICANT: Allen, Larry
; APPLICANT: Aikens, John
; APPLICANT: Fonstein, Michael
; APPLICANT: Vonstein, Veronika
; APPLICANT: Demirjian, David
; APPLICANT: Casadaban, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; FILE REFERENCE: 95-963-H
; CURRENT APPLICATION NUMBER: US/09/058,260B
; CURRENT FILING DATE: 1999-04-10
; EARLIER APPLICATION NUMBER: 60/001,995
; EARLIER FILING DATE: 1996-08-07
; EARLIER APPLICATION NUMBER: 60/009,704
; EARLIER FILING DATE: 1996-01-11
; EARLIER APPLICATION NUMBER: 08/781,802
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 08/827,810
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27

Query Match          70.4%; Score 16.2; DB 4; Length 2345;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattctgcttcgctgtttt 21
    ||||| ||| ||| ||||| |||||
Db 2124 gaatccctccatcgctgtttt 2144

RESULT 3
US-09-058-260-27
; Sequence 27, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
; APPLICANT: Allen, Larry
; APPLICANT: Aikens, John
; APPLICANT: Fonstein, Michael
; APPLICANT: Vonstein, Veronika
; APPLICANT: Demirjian, David
; APPLICANT: Casadaban, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; FILE REFERENCE: 95-963-H
; CURRENT APPLICATION NUMBER: US/09/058,260B
; CURRENT FILING DATE: 1999-04-10
; EARLIER APPLICATION NUMBER: 60/001,995
; EARLIER FILING DATE: 1996-08-07
; EARLIER APPLICATION NUMBER: 60/009,704
; EARLIER FILING DATE: 1996-01-11
; EARLIER APPLICATION NUMBER: 60/019,580
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: 08/694,078
; EARLIER FILING DATE: 1996-08-08
; EARLIER APPLICATION NUMBER: 08/781,802
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 08/827,810
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27

;
; LENGTH: 2405
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
; OTHER INFORMATION: gene from bacteria E017
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (258)..(1766)
; US-09-058-260-27

Query Match          70.4%; Score 16.2; DB 4; Length 2405;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattctgcttcgctgtttt 21
    ||||| ||| ||| ||||| |||||
Db 2200 gaatccctccatcgctgtttt 2220

RESULT 4
US-09-058-260-1
; Sequence 1, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
; APPLICANT: Allen, Larry
; APPLICANT: Aikens, John
; APPLICANT: Fonstein, Michael
; APPLICANT: Vonstein, Veronika
; APPLICANT: Demirjian, David
; APPLICANT: Casadaban, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; FILE REFERENCE: 95-963-H
; CURRENT APPLICATION NUMBER: US/09/058,260B
; CURRENT FILING DATE: 1999-04-10
; EARLIER APPLICATION NUMBER: 60/001,995
; EARLIER FILING DATE: 1996-08-07
; EARLIER APPLICATION NUMBER: 60/009,704
; EARLIER FILING DATE: 1996-01-11
; EARLIER APPLICATION NUMBER: 60/019,580
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: 08/694,078
; EARLIER FILING DATE: 1996-08-08
; EARLIER APPLICATION NUMBER: 08/781,802
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 08/827,810
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3513
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
; OTHER INFORMATION: gene from bacteria E001
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1182)..(2690)
; US-09-058-260-1

Query Match          70.4%; Score 16.2; DB 4; Length 3513;
Best Local Similarity 85.7%; Pred. No. 86;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattctgcttcgctgtttt 21
    ||||| ||| ||| ||||| |||||
Db 3125 gaatccctccatcgctgtttt 3145

RESULT 5
```


US-08-781-802-9
; Sequence 9, Application US/08781802
; Patent No. 5969121
; GENERAL INFORMATION:
; APPLICANT: ALLEN, Larry
; APPLICANT: AIKENS, John
; APPLICANT: FOSTEIN, Michael
; APPLICANT: VONSTEIN, Veronika
; APPLICANT: DEMIRJIAN, David
; APPLICANT: CASADABAN, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 S. Wacker Drive 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,802
; FILING DATE: 10-JAN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/694,078
; FILING DATE: 07-AUG-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/019,580
; FILING DATE: 12-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/009,704
; FILING DATE: 11-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/001,995
; FILING DATE: 01-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 95,963-E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3545 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1397..2905
; OTHER INFORMATION: /note= "E019 sequence of longest
; OTHER INFORMATION: open reading frame; upstream untranslated region not exact"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1397..2905
US-08-781-802-9

Query Match 70.4%; Score 16.2; DB 2; Length 3545;
Best Local Similarity 85.7%; Pred. No. 86;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattctgtctgtctgtttt 21

Db 3340 GAATCCCTCCATCGTGTGTTT 3360
||||| ||| | ||||| |||||
; Sequence 1, Application US/08781802
; Patent No. 5969121
; GENERAL INFORMATION:
; APPLICANT: ALLEN, Larry
; APPLICANT: AIKENS, John
; APPLICANT: FOSTEIN, Michael
; APPLICANT: VONSTEIN, Veronika
; APPLICANT: DEMIRJIAN, David
; APPLICANT: CASADABAN, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 S. Wacker Drive 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60605
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,802
; FILING DATE: 10-JAN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/694,078
; FILING DATE: 07-AUG-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/019,580
; FILING DATE: 12-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/009,704
; FILING DATE: 11-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/001,995
; FILING DATE: 01-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 95,963-E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4315 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1182..2690
; OTHER INFORMATION: /note= "E001, longest open reading
; OTHER INFORMATION: frame; other possible start codons at ATG/mets; GTG/val8;
; OTHER INFORMATION: GTG/val10; TTG/leu17"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1182..2690
US-08-781-802-1

Query Match 70.4%; Score 16.2; DB 2; Length 4315;

Best Local Similarity 85.7%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 3; Indels 3; Gaps 0;
Qy 1 gaattcgttcgctgtttt 21
||||| ||| | |||||
Db 3125 GAATCCCTCCATCGCTGTTT 3145

RESULT 7
US-08-694-078-1
; Sequence 1, Application US/08694078
; Patent No. 6218163
; GENERAL INFORMATION:
; APPLICANT: ALLEN, Larry
; APPLICANT: AIKENS, John
; APPLICANT: FONSTEIN, Michael
; APPLICANT: VONSTEIN, Veronika
; APPLICANT: DEMIRJIAN, David
; APPLICANT: CASADABAN, Malcolm
; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff, Ltd.
; STREET: 300 S. Wacker Drive 7th Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/694,078
; FILING DATE: 07-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/019,580
; FILING DATE: 12-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/009,704
; FILING DATE: 10-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/001,995
; FILING DATE: 07-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 95,963-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1182..2690
; OTHER INFORMATION: /note= "E001, longest open reading
; OTHER INFORMATION: frame; other possible start codons at ATG/met5; GTG/val8;
; OTHER INFORMATION: GTG/val10; TGG/leu17"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1182..2690
; US-08-694-078-1

Query Match 70.4%; Score 16.2; DB 4; Length 4315;

Best Local Similarity 85.7%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 3; Indels 3; Gaps 0;
Qy 1 gaattcgttcgctgtttt 21
||||| ||| | |||||
Db 3125 GAATCCCTCCATCGCTGTTT 3145

RESULT 8
US-09-008-979A-8
; Sequence 8, Application US/09008979A
; Patent No. 6080914
; GENERAL INFORMATION:
; APPLICANT: Conner, Timothy W.
; TITLE OF INVENTION: Strawberry Promoters and Genes
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carmen Rodriguez, Paralegal, Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,979A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Kammerer, Patricia A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: MOBT:058 (38-21(10613)A)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 787-1438
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 752 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-008-979A-8

Query Match 68.7%; Score 15.8; DB 3; Length 752;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 ttctcgttcgctgttttc 22
||||| ||| | |||||
Db 40 TTCTCTCTCTCTCTGTTTC 58

RESULT 9
US-09-460-618-8
; Sequence 8, Application US/09460618
; Patent No. 6235482
; GENERAL INFORMATION:
; APPLICANT: Conner, Timothy W.
; TITLE OF INVENTION: Strawberry Promoters and Genes
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carmen Rodriguez, Paralegal, Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/460,618
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/09/008,979
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kammerer, Patricia A.
;; REGISTRATION NUMBER: 29,775
;; REFERENCE/DOCKET NUMBER: M08T:058 (38-21(10613)A)
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (713) 787-1438
;; TELEFAX: (713) 789-2679
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 752 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; US-09-460-618-8

Query Match 68.7%; Score 15.8; DB 4; Length 752;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ttctctgcttcgtgttttc 22
|| ||||| |||||
DB 40 TTTCTGCTTCTGTTTC 58

RESULT 10
US-08-474-140-15
; Sequence 15, Application US/08474140
; Patent No. 5721127
; GENERAL INFORMATION:
; APPLICANT: DEWEER, PHILIPPE
; APPLICANT: AMORY, ANTOINE
; TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH
; TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS
; TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WILLIAM BRINKS HOFER GILSON & LIONE, P.C.
; STREET: 2000 K Street, N.W., Suite 200
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,140
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilhlem F. Gagliano, Esq.
; REGISTRATION NUMBER: 37,136
; REFERENCE/DOCKET NUMBER: 4121-41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 429-0625
; TELEFAX: (202) 293-1850
; TELEX: 650 383-5605
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 752 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: nucleic acid
; US-09-460-618-8

;; LENGTH: 431 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: nucleic acid
;; US-08-474-140-15

Query Match 67.8%; Score 15.6; DB 1; Length 431;
Best Local Similarity 81.8%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 aattctgcttcgctgtttcca 23
| | | | | | | | | | | | | | |
DB 66 ATTTGGCGCTTCGGTGTTC 87

RESULT 11
US-08-477-630-15
; Sequence 15, Application US/08477630
; Patent No. 5721128
; GENERAL INFORMATION:
; APPLICANT: DEWEER, PHILIPPE
; APPLICANT: AMORY, ANTOINE
; TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH
; TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS
; TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WILLIAM BRINKS HOFER GILSON & LIONE, P.C.
; STREET: 2000 K Street, N.W., Suite 200
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,630
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilhlem F. Gagliano, Esq.
; REGISTRATION NUMBER: 37,136
; REFERENCE/DOCKET NUMBER: 4121-42
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 429-0625
; TELEFAX: (202) 293-1850
; TELEX: 650 383-5605
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: nucleic acid
; US-08-477-630-15

Query Match 67.8%; Score 15.6; DB 1; Length 431;
Best Local Similarity 81.8%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 aattctgcttcgctgtttcca 23
| | | | | | | | | | | | | | |
DB 66 ATTTGGCGCTTCGGTGTTC 87

RESULT 12
US-08-472-293-15

; Sequence 15, Application US/08472293
; Patent No. 5731174
; GENERAL INFORMATION:
; APPLICANT: DEWEER, PHILIPPE
; APPLICANT: AMORY, ANTOINE
; TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH
; TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS
; TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WILLIAN BRINKS HOFER GILSON & LIONE, P.C.
; STREET: 2000 K Street, N.W., Suite 200
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,293
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilhem F. Gadiano, Esq.
; REGISTRATION NUMBER: 37,136
; REFERENCE/DOCKET NUMBER: 4121-44
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 429-0625
; TELEFAX: (202) 293-1850
; TELEX: 650 383-5605
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: nucleic acid
; US-08-472-293-15

Query Match 67.8%; Score 15.6; DB 1; Length 431;
Best Local Similarity 81.8%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 aattcctgcttcgctgtttca 23
| || | ||||| |||||
Db 66 ATTTCGGCTTCGGTGTTC 87

RESULT 13
US-08-474-545-15
; Sequence 15, Application US/08474545
; Patent No. 5736375
; GENERAL INFORMATION:
; APPLICANT: DEWEER, PHILIPPE
; APPLICANT: AMORY, ANTOINE
; TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH
; TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS
; TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WILLIAN BRINKS HOFER GILSON & LIONE, P.C.
; STREET: 2000 K Street, N.W., Suite 200
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,545
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilhem F. Gadiano, Esq.
; REGISTRATION NUMBER: 37,136
; REFERENCE/DOCKET NUMBER: 4121-43
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 429-0625
; TELEFAX: (202) 293-1850
; TELEX: 650 383-5605
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: nucleic acid
; US-08-474-545-15

Query Match 67.8%; Score 15.6; DB 1; Length 431;
Best Local Similarity 81.8%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 aattcctgcttcgctgtttca 23
| || | ||||| |||||
Db 66 ATTTCGGCTTCGGTGTTC 87

RESULT 14
US-08-478-341-15
; Sequence 15, Application US/08478341
; Patent No. 5817498
; GENERAL INFORMATION:
; APPLICANT: DEWEER, PHILIPPE
; APPLICANT: AMORY, ANTOINE
; TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH
; TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS
; TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WILLIAN BRINKS HOFER GILSON & LIONE, P.C.
; STREET: 2000 K Street, N.W., Suite 200
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,341
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilhem F. Gadiano, Esq.
; REGISTRATION NUMBER: 37,136
; REFERENCE/DOCKET NUMBER: 4121-45
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 429-0625
; TELEFAX: (202) 293-1850
; TELEX: 650 383-5605
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

Query Match 67.88; Score 15.6; DB 3; Length 431;

THIS PAGE BLANK (USPTO)

IntelliGenetics

FastDB - Fast Pairwise Comparison of Sequences Release 5.4

Results file seq2-1-1071aa.res made by bobryen on Tue 23 Jul 102 15:28:45-PDT.

Query sequence being compared:US-09-235-416-2 (1-1071)
 Number of sequences searched: 1
 Number of scores above cutoff: 1

Results of the initial comparison of US-09-235-416-2 (1-1071) with:
File : US09235416A.pep

	100-	90-	80-	70-	60-	50-	40-	30-	20-	10-	0
N											
U											
M											
B											
E											
R											
O											
F											
S											
E											
Q											
U											
E											
N											
C											
E											
S											

	100-	90-	80-	70-	60-	50-	40-	30-	20-	10-	0
N											
U											
M											
B											
E											
R											
O											
F											
S											
E											
Q											
U											
E											
N											
C											
E											
S											

	100-	90-	80-	70-	60-	50-	40-	30-	20-	10-	0
N											
U											
M											
B											
E											
R											
O											
F											
S											
E											
Q											
U											
E											
N											
C											
E											
S											

	100-	90-	8
--	------	-----	---

PARAMETERS

	PAM-150	K-tuple
Similarity matrix		1
Threshold level of sim.	16%	
Translation frame	6	
Mismatch penalty		1
Gap penalty	5.00	Joining penalty
Gap size penalty	0.05	Window size
Cutoff score	1	
Randomization group	0	

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	357	0	0.00

Times:	CPU	Total Elapsed
	00:00:00.00	00:00:00.00

Number of residues:	784
Number of sequences searched:	1
Number of scores above cutoff:	1

The scores below are sorted by initial score. Significance is calculated based on initial s

A 100% similar sequence to the query sequence was found:

Sequence Name	Description	Length	Score	Init. Opt.
1. US-09-235-416-1	Sequence 1, Application US	784	357	357 0.00 1

1. US-09-235-416-2 (1-1071)
US-09-235-416-1 Sequence 1, Application US/09235416A

Initial Score	=	357	Optimized Score	=	357	Significance	=	0.00
Residue Identity	=	100%	Matches	=	357	Mismatches	=	0
Gaps	=	0	Conservative Substitutions	=	0			
Translation Frame	=	1						

[illegible]

THIS PAGE BLANK (USP1U)

IntelliGenetics

FastDB - Fast Pairwise Comparison of Sequences Release 5.4

Results file seq2-1327-1803-aa.res made by bobryen on Tue 23 Jul 102 15:29:45-PDT.

Query sequence being compared: US-09-235-416-2 (1327-1803)
 Number of sequences searched: 1
 Number of scores above cutoff: 1

Results of the initial comparison of US-09-235-416-2 (1327-1803) with:
File : US09235416A.pep

	SCORE	STDEV
N	18	0
U	35	1
M	53	1
B	71	1
E	88	1
R	106	1
R	124	1
O	141	1
F	159	1
S		
E		
Q		
U		
E		
N		
C		
C		
E		
S		

PARAMETERS

Similarity matrix	PAM-150	1
Threshold level of sim.	16%	1
Translation frame	6	20
Mismatch penalty	1	477
Gap penalty	5.00	Joining penalty
Gap size penalty	0.05	Window size
Cutoff score	1	
Randomization group	0	

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	159	0	0.00

Times:	CPU	Total Elapsed
	00:00:00.00	00:00:00.00

Number of residues:	784
Number of sequences searched:	1
Number of scores above cutoff:	1

The scores below are sorted by initial score. Significance is calculated based on initial s

A 100% similar sequence to the query sequence was found:

Sequence Name	Description	Length	Score	Sig.	Frame	Init.	Opt.
1. US-09-235-416-1	Sequence 1, Application US	784	159	159	0.00	1	

1. US-09-235-416-2 (1327-1803)
US-09-235-416-1 Sequence 1, Application US/09235416A

Initial Score	=	159	Optimized Score	=	159	Significance	=	0.00
Residue Identity	=	100%	Matches	=	159	Mismatches	=	0
Gaps	=		Conservative Substitutions	=			=	0
Translation Frame	=	1						

[illegible]

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

[illegible]

2. US-09-235-416-2 (1-1071)
US-09-235-416-3 Sequence 3, Application US/09235416A

Initial Score	=	21	Optimized Score	=	21	Significance	=	-0.39
Residue Identity	=	100%	Matches	=	21	Mismatches	=	0
Gaps	=	0	Conservative Substitutions	=	0		=	0

X 10 X 30 40
ATGTCGGCGGTGGAATATCATAGGTGTCGGGTACG
|||||
ATGTCGGCGGTGGAATATC
X 10 20 X

3. US-09-235-416-2 (1-1071)
US-09-235-416-6 Sequence 6, Application US/09235416A

Initial Score	=	17	Optimized Score	=	17	Significance	=	-0.40
Residue Identity	=	40%	Matches	=	17	Mismatches	=	13
Gaps	=	0	Conservative Substitutions	=	0		=	0

800 810 820 830 840 X 850 860
CGTCCACGGAGCTACTGAGCGGCACATGAGGAGGTGGAGAGATCAACGCCGTCACTTTCCACCCTAGG

||||| | | ||||| ||
GGCGGAATTCTCDGANCDCGCVARRTCNAC
X 10 20 30

4. US-09-235-416-2 (1-1071)
US-09-235-416-5 Sequence 5, Application US/09235416A

[illegible]

5. US-09-235-416-2 (1-1071)
US-09-235-416-7 Sequence 7, Application US/09235416A

Initial Score	=	12	Optimized Score	=	15	Significance	=	-0.41
Residue Identity	=	36%	Matches	=	15	Mismatches	=	15
Gaps	=	0	Conservative Substitutions	=	0		=	0

480 490 X 500 510 530 540
 GAATTTACAAATGAACGAGTGGACACTGTGTAATCGTCGCAAGAGGGGAATCTCAAGGTCGAGAC
 ||||| ||| | |||||
 GCGCGAATCTCTCDCTNCCDGCVARRTCNAC
 X 10 20 30

6. US-09-235-416-2 (1-1071)
US-09-235-416-4 Sequence 4, Application US/09235416A

```
Initial Score      = 10  Optimized Score = 11  Significance = -0.42
Residue Identity  = 58%  Matches      = 14  Mismatches  = 9
Caps             = 1    Conservative Substitutions = 0
```

290 300 X 310 320 X 330 340
AATGCATCAAGGGTTATAACAATTGTATCTTCGGCTACGGTCAGACCGTTGGCGCAAGTCCTC


```
60      70      80      90     100     110     120
ACCGAGGAATCCACAAGGACGAGCGCGCTCGAGGAGCTGGGTATCAGCATCGAAAAGGGCTTTG
||| ||| ||| ||| ||| |||
GC CGGAATTC TCTCDCTNCCDGCVARRTCAC
X      10      20      30
```

3. US-09-235-416-2 (1327-1803)

US-09-235-416-6 Sequence 6, Application US/09235416A

```
Initial Score      = 8  Optimized Score = 15  Significance = -0.40
Residue Identity  = 33% Matches = 15  Mismatches = 15
Gaps              = 0  Conservative Substitutions = 0
```

```
80      90      100     110     120     130     140
AACGAGAAGCGCGCTCGAGGAGTGGGTATCAGCATCGAAAGGGCTTTGTCGCCCTTACCACCTCAA
||| ||| ||| ||| ||| |||
GC CGGAATTC TCTCDGANCDCGVARRTCAC
X      10      20      30
```

4. US-09-235-416-2 (1327-1803)

US-09-235-416-3 Sequence 3, Application US/09235416A

```
Initial Score      = 6  Optimized Score = 9  Significance = -0.41
Residue Identity  = 42% Matches = 9  Mismatches = 12
Gaps              = 0  Conservative Substitutions = 0
```

```
40      50      60      70      80      90
CAGACCTGGGAGAGAGCTGCCCAAGCCGAGGAATCCACAGGACGAGAGCGCGC
||| ||| ||| ||| |||
ATGTCGGGCGGTGGAATATC
X      10      20
```

5. US-09-235-416-2 (1327-1803)

US-09-235-416-4 Sequence 4, Application US/09235416A

```
Initial Score      = 5  Optimized Score = 12  Significance = -0.42
Residue Identity  = 52% Matches = 12  Mismatches = 11
Gaps              = 0  Conservative Substitutions = 0
```

```
250     260     270     280     290     300
TCACCAAGATACACAAGCGGAATTCGTCTGACGCTTCGAGATCCTGAAGAACACCTGTA
||| ||| ||| ||| |||
GAATTCCTGCTCGCTGTTTCA
X      10      20      30
```


$$\begin{array}{ccc} \vee & \bigcirc & \vee \\ & \text{---} & \\ \bigcirc & & \bigcirc \\ & \text{---} & \\ \wedge & \bigcirc & \wedge \end{array}$$

FastDB - Fast Pairwise Comparison of Sequences Release 5.4

Results file us-09-235-416-2-1804-2352.res made by bobrven on Tue 23 Jul 102 15:25:18-PR

Query sequence being compared:US-09-235-416-2 (1804-2352)

Number of sequences searched:	6
Number of scores above cutoff:	6

Results of the initial comparison of US-09-235-416-2 (1804-2352) with:
File : US09235416A.seq

[illegible]

1. US-09-235-416-2	Sequence 2, Application US	2352	549	549	2.04	0
--------------------	----------------------------	------	-----	-----	------	---

The list of other best scores is:

Sequence Name	Description	Length	Score	Init. Opt.	Sig. Frame
2. US-09-235-416-7	Sequence 7, Application US	30	15	15	-0.38 0
3. US-09-235-416-4	Sequence 4, Application US	23	9	11	-0.40 0
4. US-09-235-416-3	Sequence 3, Application US	21	7	9	-0.41 0

1. US-09-235-416-2 (1804-2352)

US-09-235-416-2 Sequence 2, Application US/09235416A

Initial Score	=	549	Optimized Score	=	549	Significance	=	2.04
Residue Identity	=	100%	Matches	=	549	Mismatches	=	0
Gaps	=	0	Conservative Substitutions	=	0			

X 10 20 30 40 50
 CAAGACAATCCCTTGCTGCCAATCTGTACCAACAGTCAGTTGGGTTCGG
 AAGAACTCGTCGGGACGGCAGAGAAATCTTGCTGCCAATCTGTACCAACAGTCAGTTGGGTTCGG
 1790 1800 X 1810 1820 1830 1840 1850
 60 70 80 90 100 110 120
 CTGCTCCAGGCTGTCAGGACCGGACACTGAGCAAGCGGGTTCGGATCGGAGCGGCAATCTCGCTCAGATT
 CTGCTCCAGGCTGTCAGGACCGGACACTGAGCAAGCGGGTTCGGATCGGAGCGGCAATCTCGCTCAGATT
 1860 1870 1880 1890 1900 1910 1920
 130 140 150 160 170 180 190
 CTCCTTTGCGGCACTTCTGTGGAAGGATAGCAGCTGTCTATGCTCGCAGGCAAGCTCTAGCGCGATCC
 CTCCTTTGCGGCACTTCTGTGGAAGGATAGCAGCTGTCTATGCTCGCAGGCAAGCTCTAGCGCGATCC
 1930 1940 1950 1960 1970 1980 1990
 200 210 220 230 240 250 260
 TAGGGTTGGATCAGAAGATCTCTCATGTACACATGACGAGTTGGATGCATTATTGACCATTCTTCAGAAG
 TAGGGTTGGATCAGAAGATCTCTCATGTACACATGACGAGTTGGATGCATTATTGACCATTCTTCAGAAG
 2070 2080 2090 2100 2110 2120 2130 2140 2150
 270 280 290 300 310 320 330 340
 CGCGGCACTGCTGTGGCTGGTGGAGCAACGCAAGATAGCATTCGACAGTTCTGTTCCGTCGCCCTCG
 CGCGGCACTGCTGTGGCTGGTGGAGCAACGCAAGATAGCATTCGACAGTTCTGTTCCGTCGCCCTCG
 350 360 370 380 390 400 410
 ACAAAATACATGTCCATGGAACCATTGATAATTTCTCGTCGATACCGGCATTACTATCGGGTAGCCCTC
 ACAAAATACATGTCCATGGAACCATTGATAATTTCTCGTCGATACCGGCATTACTATCGGGTAGCCCTC
 420 430 440 450 460 470 480
 GTAGTGATGACGACGGTGACGCGCTGTTTTTGGTGATAGAAGTCGAAACAGATCGGCTCTAATGTTGATG
 GTAGTGATGACGACGGTGACGCGCTGTTTTTGGTGATAGAAGTCGAAACAGATCGGCTCTAATGTTGATG
 490 500 510 520 530 540 X
 TTGAGAGATTGGTCAACAGCAGGCTCAGATGGAAGAGCCCTGAAACAGGCAAGCAGGAATTC
 TTGAGAGATTGGTCAACAGCAGGCTCAGATGGAAGAGCCCTGAAACAGGCAAGCAGGAATTC
 2290 2300 2310 2320 2330 2340 2350 X
 2270 2280 2290 2300 2310 2320 2330 2340 2350 X

2. US-09-235-416-2 (1804-2352)

US-09-235-416-7 Sequence 7, Application US/09235416A

Initial score	15	Optimized score	15	Significance
				-0.38

Residue Identity = 36% Matches = 15 Mismatches = 15
Gaps = 0 Conservative Substitutions = 0

250 260 270 280 290 X 310
TATTTGACGATGTTACAGAAAGCGCGGCTGCTGGCTGTCGAAGACAACGAAGATAGCGATTTC
||||| | | | | |
GGCGGAATTCCTCDCTNCCDCSVARRTCNAC
X 10 20 30

3. US-09-235-416-2 (1804-2352)

US-09-235-416-4 Sequence 4, Application US/09235416A

Initial Score = 9 Optimized Score = 11 Significance = -0.40
Residue Identity = 47% Matches = 11 Mismatches = 12
Gaps = 0 Conservative Substitutions = 0

410 420 X 430 440 X 450 460
GGGTACCCCTCGTAGTGATGACGACGCGGTGAGCGGCTGTTTTTTGGTGATAGAAAGTCGAAACA
|| |||||
GAATTCCTGCTTCGCTGTTTTCA
X 10 20 X

4. US-09-235-416-2 (1804-2352)

US-09-235-416-3 Sequence 3, Application US/09235416A

Initial Score = 7 Optimized Score = 9 Significance = -0.41
Residue Identity = 42% Matches = 9 Mismatches = 12
Gaps = 0 Conservative Substitutions = 0

120 130 140 150 X 170
TCAGATTCTCCTTTGCCGCACCTTCGTGGAAGGATAGCGACTGGTTCTATGCTCGCAGGG
| | | | |
ATGTCGGGCGGTGGAATATC
X 10 20

> O <
OI IO IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file us-09-235-416-2-inv-1327-1803.res made by bobryen on Tue 23 Jul 102 15:26:5

Query sequence being compared: US-09-235-416-2' (1327-1803)
Number of sequences searched: 6
Number of scores above cutoff: 6

Results of the initial comparison of US-09-235-416-2' (1327-1803) with:
File : US09235416A.seq

complement

100-

N -
U 50-
M -
B -
E -
R -
O -
F 10-
S -
E 5-
Q -
U -
E -
N -
C -
E -
S 0-

----- * * * * *
SCORE 0 4 8 11 15 19 23 26 30 34
STDEV

PARAMETERS

Similarity matrix Unitary K-tuple 4
Mismatch penalty 1 Joining penalty 30
Gap penalty 5.00 Window size 477
Gap size penalty 0.33
Cutoff score 1
Randomization group 0

SEARCH STATISTICS

Scores: Mean 14 Median 10 Standard Deviation 10.14
Times: CPU 00:00:00.00 Total Elapsed 00:00:00.00
Number of residues: 2486
Number of sequences searched: 6
Number of scores above cutoff: 6

The scores below are sorted by initial score.
Significance is calculated based on initial score.
A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Init. Opt.
1. US-09-235-416-2	Sequence 2, Application US	2352	34	179
2. US-09-235-416-7	Sequence 7, Application US	30	14	17
3. US-09-235-416-6	Sequence 6, Application US	30	12	19
4. US-09-235-416-5	Sequence 5, Application US	30	9	13
5. US-09-235-416-4	Sequence 4, Application US	23	8	12
6. US-09-235-416-3	Sequence 3, Application US	21	7	10

1. US-09-235-416-2' (1327-1803)
US-09-235-416-2 Sequence 2, Application US/09235416A

Initial Score = 34 Optimized Score = 179 Significance = 1.97
Residue Identity = 41% Matches = 206 Mismatches = 265
Gaps = 28 Conservative Substitutions = 0

GTACTGACGTGGCTTCTGAAGGACTCTTGGGAGCGCAACTCGATGACCCGCATGATTGCGGCATTTTCGCCT
950 960 X 970 980 990 1000 1010
60 70 80 90 100 110 120
CCCAAGGAGTCTTTCAGAACCCAGCTAGTACG---GAATCTCGGTAAAGTACTAATCTTCTTCTGT
111 111 111 111 111 111 111
GCTGATTAATTTGAAGAGACTCTCAGTACCTTCGATATGCGGACTCTCGGAGCGAATCAAGAACCAC
1020 1030 1040 1050 1060 1070 1080
130 140 150 160 170 180 190
TTTTCGAGGAGATATCCGCTAGCGCTGCAATCACACG---ACCTAGGTGCGAAAGTGGAGCGGTTCATCTCT
111 111 111 111 111 111 111
GCAGTGGTCAATGAGACCCGCAACCGCGGATGATCCGCGAGTTGAGGAGGAACTCGCGCAGCTGAGGAGC
1090 1100 1110 1120 1130-- 1140 1150
GCACCTCTTTCAGTCCGCTCCAGTACGTTCCGCTGAGTTCGCTCAGAACCCGCAATC-TAGCAG
200 210 220 230 240 250 260
AACTCCAGAGCAGTGGTGGAGGTGGAGGTGGTGCAGAGGTTCTCGCGGCGCAGTGGAGGAAATCGTACCCG
1160 1170 1180 1190 1200 1210 1220 1230
ACTGA-----TGTCGCAACCTCTCTGTCCATTTGTCTTTCATCATCCACCTTTTGGGTCAAGTTC
111 111 111 111 111 111 111
CCGACACCCGCTCGAAGAACAAATCGTTCGATTTCAG---CAGCCGGATCGACAGTCAAGAAATGAG
1240 1250 1260 1270 1280 1290
330 340 350 360 370 380 390
AAAGTGAAGACGCGTGGGATCGACGTGATCTCTCTTCATGTTTGGCGGCAACCTTCTGCTGTTTG---
111 111 111 111 111 111 111
CAAG-GCAGAAATCGT-GGAGCAACTGAACCAAGTGCAGAGTCTATCGGGATCTCAATCACACCTGGGAA
1300 1310 1320 1330 1340 1350 1360
400 410 420 430 440 450 460
-----TTGCGCTCATCCATGAGATTTTCGATTTCTTGGATGATCGACACCTGCGCAAGTCTCTC
111 111 111 111 111 111 111
GAGAGCTGCCAAGACCGGAAATCCACAGGAGACGAGCGGCTCGAGGAGTGGGTATCAGCATC
1370 1380 1390 1400 1410 1420 1430 1440
470 X
CAGGTAGGGCGCGGT
111 111
GAAGGGCTTTGTGGCCCTTACCACTCCAAAGA
1450 X 1460 1470

2. US-09-235-416-2' (1327-1803)
US-09-235-416-7 Sequence 7, Application US/09235416A

Initial Score = 14 Optimized Score = 17 Significance = 0.00

Residue Identity = 43% Matches = 17 Mismatches = 13
Gaps = 0 Conservative Substitutions = 0

210 220 230 240 250 270
GCTCCAGTAGCTCCGGTGGAGTGTCTCGTCAGAACCCGCCAAATCTACGAGACTGATCTTCGCAACCT
||||| ||| ||||| |||
GCGGGAATCTCDDTNCDCGCVARRTCNAC
X 10 20 30

3. US-09-235-416-2' (1327-1803)
US-09-235-416-6 Sequence 6, Application US/09235416A

Initial Score = 12 Optimized Score = 19 Significance = -0.20
Residue Identity = 50% Matches = 19 Mismatches = 11
Gaps = 0 Conservative Substitutions = 0

210 220 230 240 250 270
GCTCCAGTAGCTCCGGTGGAGTGTCTCGTCAGAACCCGCCAAATCTACGAGACTGATCTTCGCAACCT
||||| ||| ||||| |||
GCGGGAATCTCDDTNCDCGCVARRTCNAC
X 10 20 30

4. US-09-235-416-2' (1327-1803)
US-09-235-416-5 Sequence 5, Application US/09235416A

Initial Score = 9 Optimized Score = 13 Significance = -0.49
Residue Identity = 30% Matches = 13 Mismatches = 17
Gaps = 0 Conservative Substitutions = 0

240 250 260 270 280 290 300
CGCTCAGAACCCGCCAAATCTACGAGACTGATCTTCGCAACCTTCTCTGTGTCATTTGGTCTCTTCAT
||| ||||| |||
GCGCGGATCCATYTYTGCHTAYGGNCARAC
X 10 20 30

5. US-09-235-416-2' (1327-1803)
US-09-235-416-4 Sequence 4, Application US/09235416A

Initial Score = 8 Optimized Score = 12 Significance = -0.59
Residue Identity = 52% Matches = 12 Mismatches = 11
Gaps = 0 Conservative Substitutions = 0

250 260 270 280 290 300 310
GCCAAATCTACGAGACTGATCTTCGCAACCTTCTCTGTGTCATTTGGTCTCTTCATCATGC
||| ||||| |||
GAATTCCTGCTCGCTGTTTCA
X 10 20 X

6. US-09-235-416-2' (1327-1803)
US-09-235-416-3 Sequence 3, Application US/09235416A

Initial Score = 7 Optimized Score = 10 Significance = -0.69
Residue Identity = 47% Matches = 10 Mismatches = 11
Gaps = 0 Conservative Substitutions = 0

200 210 220 230 240 250
CCCTCCTTCAGTCGGCTCCAGTAGTCCGGTGGAGCTTGTCTCGCTCAGAACCCGCCAAAT
||||| |||
ATGTCGGCGGTGGAAATATC
X 10 20


```
2. US-09-235-416-2' (1804-2352)
   US-09-235-416-4 Sequence 4, Application US/09235416A
Initial Score = 10 Optimized Score = 13 Significance = -0.29
Residue Identity = 56% Matches = 13 Mismatches = 10
Gaps = 0 Conservative Substitutions = 0

290      300      310      320      330      340
ACTCGAGATCTTGAATAGTCTTCCTGCTCGCATAGTTGGAGCATTCTTGTCAAAGGAC
      ||||| |||||
      GAATTCCTGCTTCGCTGTTTCA
      X      10      20      X

3. US-09-235-416-2' (1804-2352)
   US-09-235-416-6 Sequence 6, Application US/09235416A
Initial Score = 10 Optimized Score = 15 Significance = -0.29
Residue Identity = 36% Matches = 15 Mismatches = 15
Gaps = 0 Conservative Substitutions = 0

10      20      30      40      X      60
CCACGGGTGTTCTCGGACCTTCAGATTCCCTTTCTCGACGATTACACCAAGTCTCGCACTCGTTTCATT
      ||| || ||| ||| ||| ||| |||
      GCGCGAATTCTCDGANCDCGCVARRTCNAC
      X      10      20      30

4. US-09-235-416-2' (1804-2352)
   US-09-235-416-7 Sequence 7, Application US/09235416A
Initial Score = 9 Optimized Score = 16 Significance = -0.36
Residue Identity = 36% Matches = 16 Mismatches = 14
Gaps = 0 Conservative Substitutions = 0

390      400      X      410      420      430      X      440      450
ATAATAGTTTGGCCACTTTTACGGCCCTTCTCTCGCACCCGAGGGGAGGATGCTTTGATTTTC
      |||| ||||| |||||
      GCGCGAATTCTCDCTNCCDGCVARRTCNAC
      X      10      20      30

5. US-09-235-416-2' (1804-2352)
   US-09-235-416-3 Sequence 3, Application US/09235416A
Initial Score = 7 Optimized Score = 11 Significance = -0.51
Residue Identity = 52% Matches = 11 Mismatches = 10
Gaps = 0 Conservative Substitutions = 0

470      480      490      500      X      520      530
CATTTTGGCCACGGTCGATTTCCTCGGGCGTTGAACGGGCGGTACCCGCCACCACCACTTGA
      ||||| |||
      ATGTCGGCGGTGGAATATC
      X      10      20
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 04:39:12 ; Search time 2969.55 Seconds
(without alignments)
104.538 Million cell updates/sec

Title: US-09-235-416-4
Perfect score: 23
Sequence: 1 gaattcgtcttcgctgtttca 23

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18.8	81.7	396	12	AZ390315	AZ390315 IM0151C07
2	18.8	81.7	747	12	AZ804429	AZ804429 2M0065H14
3	18.8	81.7	882	12	AZ685448	AZ685448 ENTJY17MF
4	18.2	79.1	178	9	AA253238	AA253238 z553f07.r
5	18.2	79.1	224	9	AW844301	AW844301 RC4-CN005
6	18.2	79.1	280	10	T06084	T06084 EST03973 Fe
7	18.2	79.1	281	9	AA378449	AA378449 EST91145
8	18.2	79.1	293	9	AW844294	AW844294 RC4-CN005
9	18.2	79.1	303	9	AA355871	AA355871 ESP64335
10	18.2	79.1	306	9	AA411640	AA411640 zyl5c03.r
11	18.2	79.1	322	10	W25620	W25620 zbl7le02.r1
12	18.2	79.1	328	10	BF845084	BF845084 CM1-HT114
13	18.2	79.1	337	12	AQ260557	AQ260557 CITBI-E1-
14	18.2	79.1	356	10	BI033412	BI033412 PM1-NN120
15	18.2	79.1	382	9	AV721381	AV721381 AV721381
16	18.2	79.1	411	10	BF955757	BF955757 WR2-GN012
17	18.2	79.1	413	12	AQ333260	AQ333260 HS-5009_A

c 18	18.2	79.1	446	10	W89640	W89640 mf75c03.r1
19	18.2	79.1	447	9	AW844305	AW844305 RC4-CN005
20	18.2	79.1	459	10	R56708	R56708 Y104h08.r1
21	18.2	79.1	465	10	BF028620	BF028620 601763846
22	18.2	79.1	468	9	AA773267	AA773267 af82f03.r
23	18.2	79.1	488	10	H94360	H94360 yw54c08.r1
c 24	18.2	79.1	473	10	BE764542	BE764542 PM4-NT007
25	18.2	79.1	547	10	W06924	W06924 za90b10.r1
c 26	18.2	79.1	567	10	BG082824	BG082824 H3081806-
27	18.2	79.1	611	9	AI730423	AI730423 BNLGH1714
c 28	18.2	79.1	621	10	BM311296	BM311296 ig63406.x
29	18.2	79.1	623	10	BE540124	BE540124 601061039
c 30	18.2	79.1	638	12	BH033255	BH033255 RPCI-24-2
c 31	18.2	79.1	644	9	AW304933	AW304933 xv97c10.x
32	18.2	79.1	660	9	AI726774	AI726774 BNLGH1652
33	18.2	79.1	663	10	BF115925	BF115925 7n74g10.x
34	18.2	79.1	669	10	BG926285	BG926285 HNC33-1-D
c 35	18.2	79.1	705	9	AW131625	AW131625 xf32c08.x
36	18.2	79.1	724	10	BG329673	BG329673 602429286
c 37	18.2	79.1	729	10	BF096178	BF096178 MR3-UT009
c 38	18.2	79.1	743	9	AA027726	AA027726 wv75f02.x
c 39	18.2	79.1	756	9	AA206424	AA206424 zq58f08.s
40	18.2	79.1	775	10	BI858782	BI858782 603386385
41	18.2	79.1	780	10	BI335414	BI335414 602937992
42	18.2	79.1	812	9	AJ396274	AJ396274 AJ396274
c 43	18.2	79.1	855	9	AL552868	AL552868 AL552868
c 44	18.2	79.1	922	10	BEG12931	BEG12931 601451871
c 45	18.2	79.1	972	9	AL571242	AL571242 AL571242

ALIGNMENTS

RESULT 1
AZ390315 396 bp DNA linear GSS 03-OCT-2000
LOCUS IM0151C07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0151C07 R, DNA sequence.

ACCESSION AZ390315
VERSION AZ390315
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT

UNIVERSITY

ADDRESS

TELEPHONE

FAX

EMAIL

INSERT LENGTH

PLATE

SEQ PRIMER

CLASS

High quality sequence stop: 396.

Location/Qualifiers

1. 396

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0151C07"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

```

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      54 a   75 c   85 g   182 t
ORIGIN

```

```

Query Match      81.7%; Score 18.8; DB 12; Length 396;
Best Local Similarity 90.9%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaattctgcttcgtgttttc 22
|||||
Db 65 GAATTCCTGCTTCTCTGTTTC 86

```

```

RESULT 2
A2804429
LOCUS      747 bp      DNA      linear      GSS 16-FEB-2001
DEFINITION      2M0065H14F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0065H14 F, DNA sequence.
ACCESSION      A2804429
VERSION      A2804429.1 GI:12956752
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 747)
AUTHORS      Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D.,Weiss,R.
TITLE      Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL      Unpublished (2000)
COMMENT      Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0065 row: H column: 14
Seq primer: CTTTGTAAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 747.
Location/Qualifiers
1..747
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0065H14"

```

```

/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      167 a   154 c   188 g   238 t
ORIGIN

```

```

Query Match      81.7%; Score 18.8; DB 12; Length 747;
Best Local Similarity 90.9%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aattctgcttcgtgttttc 23
|||||
Db 573 ATTTCTGCTTCGCTGTTTAA 594

```

```

RESULT 3
A2685448
LOCUS      882 bp      DNA      linear      GSS 14-DEC-2000
DEFINITION      ENTJY17TF Entamoeba histolytica sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION      A2685448
VERSION      A2685448.1 GI:11822594
KEYWORDS      GSS.
SOURCE      Entamoeba histolytica.
ORGANISM      Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE      1 (bases 1 to 882)
AUTHORS      Loftus,B., Van Aken,S. and Fraser,C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
JOURNAL      Unpublished (2000)
COMMENT      Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjl@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 3
High quality sequence stop: 766.
Location/Qualifiers
1..882
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHO51; Site:1; Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica"

```

```

FEATURES
source

```

using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In *Genome Sequencing: A Practical Approach*, eds. M. Vaudin and B. Barrel, Oxford University Press, 1999)."

BASE COUNT 298 a 153 c 126 g 305 t

Query Match 81.7%; Score 18.8; DB 12; Length 882;
Best Local Similarity 90.9%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 aattctgcttcgtctgttttca 23

Db 722 AATTCTGTCTCACTGCTTTCA 743

RESULT 4

AA253238

LOCUS AA253238 178 bp mRNA linear EST 06-AUG-1997
DEFINITION zr53f07.r1 Soares_NhMMPu_S1 Homo sapiens cDNA clone IMAGE:667141 5' similar to TR:G1063396 G1063396 RAGA. ;, mRNA sequence.

ACCESSION AA253238

VERSION AA253238.1 GI:1885404

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 178)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.

TITLE WashU-Merck EST Project 1997

JOURNAL Unpublished (1997)

COMMENT Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LML; contact the

IMAGE Consortium (info@image.lml.gov) for further information.

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Insert Length: 598 Std Error: 0.00

Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 1.

Location/Qualifiers

1. .178

source

/organism="Homo sapiens"

/db_xref="GDB:5561108"

/db_xref="taxon:9606"

/clone="IMAGE:667141"

/clone_lib="Soares_NhMMPu_S1"

/tissue_type="Pooled human melanocyte, fetal heart, and

pregnant uterus"

/lab_host="DH10B"

/note="Organ: mixed (see below); Vector: pT7m3D-Pac

(Pharmacia) with a modified polylinker; Site.1: Not I;

Site.2: Eco RI; Equal amounts of plasmid DNA from three

normalized libraries (melanocyte 2NDHM, pregnant uterus

NBHPU, and fetal heart NBHH19W) were mixed, and ss circles

were made in vitro. Following HAP purification, this DNA

was used as tracer in a subtractive hybridization

reaction. The driver was PCR-amplified cDNAs from pools of

5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 42 a 50 c 40 g 46 t

ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 178;

Best Local Similarity 87.0%; Pred. No. 6.8e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattctgcttcgtctgttttca 23

Db 4 GAATTCCAACTTCGCTGCTTTCA 26

RESULT 5

AW844301

LOCUS AW844301 224 bp mRNA linear EST 18-MAY-2000
DEFINITION RC4-CN0050-130200-012-f01 CN0050 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW844301

VERSION AW844301.1 GI:7938284

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 224)

AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-RC4-CN0050-130

200-012-f01&t3=2000-02-13&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 22

High quality sequence stop: 224.

Location/Qualifiers

1. .224

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="CN0050"

/dev_stage="Adult"

/note="Organ: colon_normal; Vector: puc18; Site.1: SmaI;

Site.2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

62 a 59 c 50 g 53 t

BASE COUNT

ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 224;

Best Local Similarity 87.0%; Pred. No. 7e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaattcctgcttcgtctgttttca 23
||||||| ||||||| |||||||
Db 144 GAATTCCAACTCGCTGCTTTCA 166

RESULT 6
T06084 EST03973 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA
LOCUS clone HFBDO22, mRNA sequence.
DEFINITION T06084
ACCESSION T06084
VERSION T06084.1 GI:317233
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 280)
AUTHORS Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
TITLE 3,400 expressed sequence tags identify diversity of transcripts
from human brain
JOURNAL Nature Genet. 4, 256-267 (1993)
MEDLINE 93364420
COMMENT Contact: Adams, MD
The Institute for Genomic Research
932 Clopper Road, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: mdadams@tigr.org
Seq primer: M13-21.

FEATURES
source
1..280
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="ATCC (inhost):82755"
/db_xref="taxon:9606"
/clone="HFBDO22"
/clone_lib="Fetal brain, Stratagene (cat#936206)"
/note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
oligo-grt + random primed cDNA synthesis; lambdaZAP-II
vector, 1.0kb average inser size."

BASE COUNT 66 a 64 c 80 t 10 others
ORIGIN
Query Match 79.1%; Score 18.2; DB 10; Length 280;
Best Local Similarity 87.0%; Pred. No. 7.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaattcctgcttcgtctgttttca 23
||||||| ||| ||||||| ||
Db 81 GAATTCCTCCTGGCTGTTTCCA 103

RESULT 7
AA378449 281 bp mRNA linear EST 21-APR-1997
LOCUS EST91145 Synovial sarcoma Homo sapiens cDNA 5' end similar to
DEFINITION guanine nucleotide-binding protein raga, mRNA sequence.
ACCESSION AA378449
VERSION AA378449.1 GI:2030768
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 281)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,

Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palancoes,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Melssner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.

Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280

COMMENT Other_ESTs: THC189208
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423

Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES
source
1..281
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="ATCC (inhost):182850"
/db_xref="taxon:9606"
/clone_lib="Synovial sarcoma"
/sex="female"
/tissue_type="synovial membrane"
/dev_stage="adult, 20 yrs"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 77 a 72 c 61 g 71 t
ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 281;
Best Local Similarity 87.0%; Pred. No. 7.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaattcctgcttcgtctgttttca 23
||||||| ||||||| |||||||
Db 229 GAATTCCAACTCGCTGCTTTCA 251

RESULT 8
AW844294 293 bp mRNA linear EST 18-MAY-2000
LOCUS RC4-CN0050-130200-012-c08 CN0050 Homo sapiens cDNA, mRNA sequence.
DEFINITION AW844294
ACCESSION AW844294
VERSION AW844294.1 GI:7938277
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 293)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.P.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
TITLE
JOURNAL
MEDLINE

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RC4-CN0050-130
200-012-c086t3-2000-02-13&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 9
High quality sequence stop: 293.

FEATURES

source
1. .293
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CN0050"
/dev_stage="Adult"
/note="Organ: colon_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
76 a 81 c 64 g 72 t

Query Match 79.1%; Score 18.2; DB 9; Length 293;
Best Local Similarity 87.0%; Pred. No. 7.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattcctgcctgcgtgtttca 23
||||||| ||||||| ||||||| |||||||

Db 158 GAATTCCAACTTCGCTGCTTCA 180

RESULT 9

AA355871 303 bp mRNA linear EST 21-APR-1997
LOCUS EST64335 Jurkat T-cells VI Homo sapiens cDNA 5' end similar to
DEFINITION guanine nucleotide-binding protein raga, mRNA sequence.
ACCESSION AA355871
VERSION AA355871.1 GI:2008190
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 303)
ADAMSON, M.D., Kerkavag, A.R., Fleischmann, R.D., Fuldner, R.A., Bult
C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White
O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,
Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald
L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodex, A.,
Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,
Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,
Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L.,
Wei, F.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon
M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and
Venter, J.C.

TITLE Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence

JOURNAL

MEDLINE 96026280
COMMENT Nature 377 (6547 Suppl), 3-174 (1995)
Other_ESTS: THCL89208
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tcdb/ngi/ngi.html)
Seq primer: M13 Reverse.

FEATURES

source
1. .303
/organism="Homo sapiens"
/db_xref="ATCC (inhost):157999"
/db_xref="taxon:9606"
/clone_lib="Jurkat T-cells VI"
/cell_type="T-lymphocyte"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"
73 a 75 c 60 g 93 t 2 others
BASE COUNT
ORIGIN
Query Match 79.1%; Score 18.2; DB 9; Length 303;
Best Local Similarity 87.0%; Pred. No. 7.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattcctgcctgcgtgtttca 23
||||||| ||||||| ||||||| |||||||

Db 54 GAATTCCAACTTCGCTGCTTCA 76

RESULT 10

AA411640 306 bp mRNA linear EST 17-MAY-1997
LOCUS zvl5c03.r1 Soares_NhMpu_S1 Homo sapiens cDNA clone IMAGE:753700 5'
DEFINITION similar to TR:G1063396 G1063396 RAGA. ;, mRNA sequence.
ACCESSION AA411640
VERSION AA411640.1 GI:2069365
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 306)
HILLIER, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie,
T., Waterston, R. and Willson, R.
WashU-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

TITLE This clone is available royalty-free through LLNL ; contact the
JOURNAL IMAGE Consortium (info@image.llnl.gov) for further information.
COMMENT 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 305.

FEATURES

source
1. .306
/organism="Homo sapiens"
/db_xref="GDB:597656"
/db_xref="taxon:9606"
/clone_lib="Soares_NhMpu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and

pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT 81 a 76 c 71 g 78 t
ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 306;
Best Local Similarity 87.0%; Pred. No. 7.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaattctgcttcgctgtttca 23
||||||| ||||||| |||||
Db 102 GAATTCCAACTTCGCTGCTTCA 124

RESULT 11
LOCUS W25620 322 bp mRNA linear EST 20-AUG-1996
DEFINITION Z671e02.rl Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
IMAGE:309050 5' similar to WF:T24Fl.1 CE02361 GTP-BINDING PROTEIN
; mRNA sequence.

ACCESSION W25620
VERSION W25620.1 GI:1303455
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 322)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasakis,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478

TITLE Contact: Wilson RK
JOURNAL Washington University School of Medicine
MEDLINE 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
COMMENT Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 738 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 234.
Location/Qualifiers
1..322

SOURCE
/organism="Homo sapiens"
/db_xref="GDB:1252463"
/db_xref="taxon:9606"
/clone="IMAGE:309050"
/clone_lib="Soares_fetal_lung_NbHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTACCACTCAAGTGGGAGCGCGCAATTTTTTTTTTTT-3'],

double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NBH19W."

BASE COUNT 86 a 84 c 70 g 77 t 5 others
ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 322;
Best Local Similarity 87.0%; Pred. No. 7.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaattctgcttcgctgtttca 23
||||||| ||||||| |||||
Db 145 GAATTCCAACTTCGCTGCTTCA 167

RESULT 12
LOCUS BF845084/c 328 bp mRNA linear EST 13-JAN-2001
DEFINITION CM1-HT1145-271200-677-d08 HT1145 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF845084
VERSION BF845084.1 GI:12201298
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 328)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM1&tl2=CM1-HT1145-
271200-677-d08&t3=2000-12-27&t4=1)
Seq primer: puc.18 forward
High quality sequence start: 8
High quality sequence stop: 327.
Location/Qualifiers
1..328

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT1145"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
79 a 75 c 85 t

BASE COUNT 79 a 75 c 85 t
ORIGIN

```

Query Match      79.1%; Score 18.2; DB 10; Length 328;
Best Local Similarity 87.0%; Pred. No. 7.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattcgtctgcgtgttttca 23
||||||| ||||||| |||||
Db 208 GAATTCCAACTTCGCTGCTTCA 186

RESULT 13
LOCUS      AQ260557      337 bp      DNA      linear      GSS 24-OCT-1998
DEFINITION CITBI-EI-2508E24.TF CITBI-EI Homo sapiens genomic clone 2508E24,
DNA sequence.
ACCESSION  AQ260557
VERSION     AQ260557.1 GI:3787081
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 337)
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
JOURNAL   Unpublished (1998)
COMMENT   Other_GSSs: CITBI-EI-2508E24.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES   source
            location/Qualifiers
            1..337
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="2508E24"
                /clone_lib="CITBI-EI"
                /sex="male"
                /cell_type="sperm"
            /note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
            Caltech Human BAC Library D"
BASE COUNT  98 a 67 c 75 g 97 t
ORIGIN

Query Match      79.1%; Score 18.2; DB 12; Length 337;
Best Local Similarity 87.0%; Pred. No. 7.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattcgtctgcgtgttttca 23
||||||| || ||||| |||||
Db 52 GAATTCCTCTCTGCTGCTTTTA 74

RESULT 14
BI033412/c
LOCUS      BI033412      356 bp      mRNA      linear      EST 14-JUN-2001
DEFINITION PM1-NN1207-150201-023-b07 NN1207 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BI033412
VERSION     BI033412.1 GI:14440038
KEYWORDS   EST.

```

```

SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 356)
AUTHORS        Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE          Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL        Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE        20202663
COMMENT        Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM1&t2=PM1-NN1207-
            150201-023-b07&t3=2001-02-15&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 16
            High quality sequence stop: 317.
            location/Qualifiers
            1..356
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="NN1207"
                /dev_stage="Adult"
                /note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI;
                Site_2: SmaI; A mini-library was made by cloning products
                derived from ORESTES PCR (U.S. Letters Patent application
                No. 196,716 - Ludwig Institute for Cancer Research)
                profiles into the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
BASE COUNT     54 a 77 c 92 g 93 t
ORIGIN

Query Match      79.1%; Score 18.2; DB 10; Length 356;
Best Local Similarity 87.0%; Pred. No. 7.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaattcgtctgcgtgttttca 23
||||||| ||||||| |||||
Db 93 GAATTCCAACTTCGCTGCTTCA 71

RESULT 15
AV721381
LOCUS      AV721381      382 bp      mRNA      linear      EST 16-OCT-2000
DEFINITION AV721381 HTB Homo sapiens cDNA clone HTBADE08 5', mRNA sequence.
ACCESSION  AV721381
VERSION     AV721381.1 GI:10818533
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 382)
AUTHORS    Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
            ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
TITLE      Homo sapiens cDNA HTB clones

```

JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1. .382
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTBADE08"
/clone_lib="HTB"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pbluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 81 a 99 c 78 g 124 t
ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 382;
Best Local Similarity 87.0%; Pred. No. 7.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 gaattcctgcttcgctgtttca 23
||||||| ||||||| |||||
Db 70 GAATTCACACTCGCTGCTTCA 92

Search completed: July 25, 2002, 04:39:19
Job time: 3157 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2002, 04:01:22 ; Search time 63.5 Seconds
(without alignments)
1371.368 Million cell updates/sec

Title: US-09-235-416-1

Perfect score: 4030

Sequence: 1 MSGGNIKVVVRFPNARE.....ELRQQQAQMEALKTAQGEF 784

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4030	100.0	784	20	Thermomyces lanugi
2	1688.5	41.4	1816	21	Human kinesin-like
3	1660.5	41.2	893	22	Human polypeptide
4	1658	41.1	1103	21	Human KLIMP protei
5	1658	41.1	1103	22	Human kinesin-like
6	1635	40.6	1773	22	Drosophila melanog
7	1430	35.5	757	22	Human diagnostic a
8	1396.5	34.7	1921	22	Drosophila melanog
9	1278.5	31.7	504	21	Gene 5 human secre
10	1276	31.7	503	21	Human secreted pro
11	1251	31.0	421	22	Human polypeptide

12	1128	28.0	1174	22	AB61704	Drosophila melanog
13	1035	25.7	1121	22	AB672021	Drosophila melanog
14	814	20.2	955	15	AAR57365	K39 polypeptide of
15	814	20.2	955	17	AAW03691	Leishmania chagasi
16	781.5	19.4	677	22	AB65183	Drosophila melanog
17	750.5	17.6	784	22	AB671112	Drosophila melanog
18	691.5	17.2	963	22	AAW78880	Human protein seq
19	683.5	17.0	979	22	AAW78864	Human protein seq
20	679	16.8	1212	22	AB671114	Drosophila melanog
21	673	16.7	1057	22	AAG67419	Amino acid sequenc
22	673	16.7	1057	22	AAB47212	Human KSP. Homo s
23	671.5	16.7	513	22	AAB47214	Human KSP-K491. S
24	671.5	16.7	575	22	AAB47215	Human KSP-S553. S
25	668	16.6	2633	22	ABG06505	Novel human diagno
26	663	16.5	2663	22	AAW39097	Human polypeptide
27	662.5	16.4	2954	20	AAW01632	Amino acid sequenc
28	651.5	16.2	975	22	AB63485	Drosophila melanog
29	649.5	16.1	975	19	AAW72746	Drosophila kinesin
30	645.5	16.0	2688	22	AAW40883	Human polypeptide
31	640	15.9	1003	22	AB61405	Drosophila melanog
32	633	15.7	1184	22	AAG67415	Amino acid sequenc
33	630	15.6	814	22	AB65317	Drosophila melanog
34	628	15.6	368	22	AAB47216	Human KSP-K368. S
35	625	15.5	411	19	AAW72745	Drosophila kinesin
36	625	15.5	441	19	AAW72744	Drosophila kinesin
37	616.5	15.3	382	22	AAB47213	Human KSP-K360. Sy
38	610.5	15.1	1066	22	AAG67418	Amino acid sequenc
39	603.5	15.0	1048	22	AB659245	Drosophila melanog
40	601	14.9	1029	22	ABG20649	Novel human diagno
41	587.5	14.6	1038	22	AAG67416	Amino acid sequenc
42	584	14.5	1518	21	AAG40075	Arabidopsis thalia
43	584	14.5	1662	21	AAG31282	Arabidopsis thalia
44	580.5	14.4	1460	21	AAG40077	Arabidopsis thalia
45	580.5	14.4	1462	21	AAG40076	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AAW06618

ID AAY06618 standard; Protein; 784 AA.

XX

AC AAY06618;

XX

DT 26-OCT-1999 (first entry)

XX

DE Thermomyces lanuginosus kinesin motor protein TL-gamma.

XX

KW TL-gamma; kinesin; motor protein; microtubule; unc-104; infection; neurodegenerative disease; Alzheimer's disease;

KW Parkinson's disease; Huntington's disease;

KW amyotrophic lateral sclerosis.

XX

OS Thermomyces lanuginosus.

XX

PN WO9937659-A1.

XX

PD 29-JUL-1999.

XX

PF 22-JAN-1999; 99WO-US01355.

XX

PR 23-JAN-1998; 98US-0072361.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Goldstein LSB, Sakowicz R;

XX

DR WPI; 1999-493950/41.

DR N-PSDB; AAX87656.

XX

PT New nucleic acid encoding microtubule motor protein, used for diagnosis of fungal infection and neurodegenerative disease

XX	Claim 5; Page 70-71; 75pp; English.
PS	This sequence represents Thermomyces lanuginosus TL-gamma, a novel
CC	ATP-dependent, plus end-directed microtubule motor protein that is
CC	a member of the unc-104 family and kinesin superfamily. The
CC	invention provides TL-gamma nucleic acids (see AAX87656), proteins
CC	and antibodies, and methods of screening for TL-gamma modulators
CC	potentially useful for treating hyphal fungal infections and
CC	diseases caused by mutated TL-gamma, e.g. neurodegeneration
CC	Involving anterograde axonal transport, such as Alzheimer's,
CC	Parkinson's or Huntington's diseases or amyotrophic lateral
CC	sclerosis. Detection of TL-gamma allows differentiation between
CC	hyphal and non-hyphal fungal infections.
XX	
SQ	Sequence 784 AA;
	Query Match 100.0%; Score 4030; DB 20; Length 784;
	Best Local Similarity 100.0%; Pred. No. 1.8e-299;
	Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY *	1 MSGGNIKVVVRPFNFNAREIDRGAKCIVRMEGNQITILTPPPGAEEKARKSGKTIMDGPk 60
Dd	
Dd	1 msggnikvvvrpfnfnaeidrgakcivrmegnqitiltpppgaekarksgktimdgpk 60
QY	61 AFAPDRSWSFDKNAPYARQEDIFODLGVPLLDNAFKGYNNCIFAYGQTCSGSXSNMG 120
Dd	
Dd	61 afapdrswsfdknapyarqedifodlglvplldnafkgynnclfaygqtcsgsxsnmg 120
QY	121 YGKEHGVIPRICODMFRINELQDKNLCTVEVSYLEIYNERNVRLDLNPSTKGNLKVRE 180
Dd	
Dd	121 ygkehgvipricodmfrinelqdknltctvevsyleinyernvrdllnpstkgnlkvre 180
QY	181 HPSTGYPVEDIAKLIVRSFOEINLMDGCKARTVAATNMNETSSRHAVFTLTQKWH 240
Dd	
Dd	181 hpstgpyvediaklvrsfgeienlmdegckartvaatnmnetssrhavftiltqkwh 240
QY	241 DEETKMDEKVAKISLVLDLAGSERATSGATCARLKEGAENRSLTIGRVIAALADMSS 300
Dd	
Dd	241 deetkmdtekvakislvdlagseratsgatgarikegaeinrsltigrviaaaladmss 300
QY	301 GKQKNQLVPYRDSVLTWLLKDSLGGNSMTAMIAISPADINFETILTSLRYADS AKRIK 360
Dd	
Dd	301 gkqknqlvpyrdsvltwllkds lggnsmtamiaaispadinfetiltslryadsakrik 360
QY	361 NHAVVNEDPNARMIRELKEELAOLASKLSQSGGGAGSGGPGVEESYPPDTPEKQIV 420
Dd	
Dd	361 nhavvnedpnarmirelkeelaqlrsklqs sggggaggsgpgveesyppdtplekqiv 420
QY	421 SIQQPDATVKMSKAIEVQLNQSEKLYRDNLQNTWEEKLAQTEIH KEREAALEEIGTSI 480
Dd	
Dd	421 siqqpdatkmskaievelnqsekllyrdlnqntweeklakeeihkereaaeleeigtsi 480
QY	481 EKGFGVPYHSKEMPHLVNLSDDPLAECLVYNIKPQOTRVGNVNDQTQAEIRLNGSKILK 540
Dd	
Dd	481 ekfgvgpyhskemphlvnlsddplaeclvynikpqotrvgnvndqtqaeirlngskilk 540
QY	541 EHCTFENVDNVTIVPNKAAVMVANGVRI DKPTRLSRGYRIILGFHFIFRNHPPEARAE 600
Dd	
Dd	541 ehctfenvdnvtivpnkkaavmvangvridkptrlsrgyriilgfhfirfnhppearae 600
QY	601 RQEQSILLRHSVNSQLSGSPAGRHDRTLSKAGSDADGDSRSDSLPHFFRGKSDWFYARR 660
Dd	
Dd	601 rqeqsllrhsvnsqsgspagrhdrtlskagsdadgdsrsdspiphfrgksdwfyarr 660
QY	661 EAASATILGLDQIKSHLTDELDALEDFDVOKARAVRGLVEDNEDSDSQSFPPVRDKYMSN 720
Dd	
Dd	661 eaasatilgl dqikshltddelaldelfdvokaravr glvedncdsqsfppvr dkymsn 720
QY	721 GTIDNFSLDTAITMPGTPSRSDDGDALEFFGDKKSKODASNVDVEELRQOQAOMEAEALKTA 780
Dd	
Dd	721 gtidnfsldta itmpgtprsd dgda leffgd kkskodasnvdveelrq qagameealkta 780

393 dmtalvgm-spssslsalsraasvsslherillfapggseeaierlketekiaaelnetw 451
456 EEKLAATEIHKEREAAELGELISIEK--GFVGPYHSHKEMPHLVNLSDDPLLAELVYNI 513
452 eeklrtealrmereallaemgvmredgdtlgvfpkktphlvnlndplmsecllyyi 511
514 KPGQTRGVNVDQAEIRLNGSKILKEHCTPEN-----VDNVVTIVPNEKAAVMVNGVR 568
512 kgdrtvrgredgerrdglvshghfkeehcvfrsdrsgseavvtlepecgadtynghk 571
569 IDKPTLRSGYRIILGDFHFRNPEEAREARQEQSLRHVSNSQLGSPAPGRHDTL 628
572 vtepsilrsgnrlimgkshvfrfthpeqarqerer----- 606
629 SKAGSDAGDSRSDSLPHFRGKDSDFWYARREAAASAILGLDOKISHLTDDELDAFDV 688
607 -----tpcaetpaepvdfwafqrellek-qgidmk--qemeqrlqeledgy 649
689 QKARAVRRGLVEDNEDSDSQSPFVRDKYMSN 720
650 rrereeatylle-qqrldyesklealqkqmds 680
RESULT 4
AAV51328
ID AAV51328 standard; Protein; 1103 AA.
AC AAV51328;
XX
XX
17-APR-2000 (first entry)
XX Human KLIMP protein.
XX
XX KLIMP; kinesin-like motor protein; cytosolic; anticonvulsant; human;
KW anti-Alzheimer; anti-Parkinsonian; antidiabetic; anti-ulcerative; cancer;
KW immunomodulatory; anti-inflammatory; anti-AIDS; antirheumatic; treatment;
KW antiarthritic; diagnosis; neurological disorder; vesicular transport.
XX
OS Homo sapiens.
XX
XX US6013454-A.
XX
XX 11-JAN-2000.
XX
XX 28-SEP-1998; 98US-0162373.
XX
XX 28-SEP-1998; 98US-0162373.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Tang YT, Corley NC, Patterson C, Guegler KJ;
XX
XX WPI; 2000-126064/11.
XX
XX N-PSDB; AAZ44744.
XX
XX Nucleic acid sequences encoding a human kinesin-like motor protein
XX (KLIMP) useful for the treatment of diseases associated with
XX inappropriate KLIMP expression such as cancers, neurological disorders
XX and disorders of vesicular transport -
XX
XX Claim 1; Fig 1A-J; 38pp; English.
XX
XX This invention describes a novel human kinesin-like motor protein
XX (KLIMP) (I) which has cytostatic, anticonvulsant, anti-Alzheimer's,
XX anti-Parkinsonian, antidiabetic, anti-ulcerative, immunomodulatory,
XX anti-inflammatory, anti-AIDS, antirheumatic and antiarthritic activity.
XX (I) and the protein it encodes may be used in the prevention, treatment
XX and diagnosis of diseases associated with inappropriate KLIMP expression
XX such as cancers, neurological disorders and disorders of vesicular
XX transport. For example, (I) and vectors containing (I) (iv) and the
XX KLIMP polypeptide may be used to treat disorders associated with
XX decreased KLIMP expression such as cancers (e.g. lymphoma, melanoma and

CC cancers of the breast lung and prostate), neurological disorders (e.g.
CC epilepsy, Alzheimer's disease and Parkinson's disease), disorders of
CC vesicular transport (e.g. diabetes mellitus/insipidus, Grave's disease
CC and gastric/duodenal ulcers), and some immune/inflammatory diseases (e.g.
CC acquired immune deficiency syndrome AIDS), rheumatoid arthritis and toxic
CC shock syndrome). This sequence represents the human KLIMP protein
CC described in the method of the invention.
XX
XX Sequence 1103 AA;
SQ
Query Match 41.1%; Score 1658; DB 21; Length 1103;
Best Local Similarity 47.3%; Pred. No. 1e-117;
Matches 353; Conservative 120; Mismatches 159; Indels 114; Gaps 16;
QY 4 GGNIKVVVRVPFNAEIDRGAKCIVRMENQNTILTPPPGAEKARKSGKTIMDGPKAPA 63
DB 3 gasvkavvrfrfnaretsqdkacvsmagnntslinp-----kqakdapsft 51
QY 64 FDRSYWSFKNA-PNYARDELFDQLGVPLLDNAFGYNNCIPAYGQTGSGSKSYMMGYG 122
DB 52 fdysywshtstedpqfasqqvvyrdigeemlllhafegynvcifayqgtgagksytmnrg 111
QY 123 K--EHGVIPRICODMERRINELQKDKNLCTVEVSYLEIYNERVRLNPFSTGNLKVRE 180
DB 112 epqggvlpqcedlfrvsenq-saqlsyvsvevsymelycervrlllnpkrgslrvie 170
QY 181 HPSTGYPVEDLAKLVVRFSQEIENLMDENKARTVAATNMNNTSSRSHAVFTLTQKWH 240
DB 171 hplpgyvqdlslavtsyadiadlmdcgnkartvaatnmnntssrsHAVFTLVfgrch 230
QY 241 DEETKMDTEKAKISLVDLAGSERATSTGATGARLEKAGEINRSITGLGVIAALADMSS 300
DB 231 dgitgidsekvsilvdlagseradsgarmglkeganinksltlgkvisaladms 290
QY 301 GKQKKQLVPYRDSVLTLLKDSLGGNSMTAMIAATSPADINFEETLSTLYADSAKRK 360
DB 291 -kkrksdfipyrdsvtllwkenlgnsrtamaalaspadinyeelstlryadrktqir 349
QY 361 NHAVNEDPNARMIRELKELAQLRSKLOSSGGGGGAGG----- 400
DB 350 cnalinedpnarlirelqeavarlrellmaqglasaleglkteegsvrgalpavapsppa 409
QY 401 -----SGGPVEESYPDDTPLEKQIVSIQPDATVTKMSKAEIVEQLNQSEKLYRDL 451
DB 410 pvspsptthngelapsfnt-esqi-----gpeeamerlqetekiaael 454
QY 452 NOTWEEKLAETEEIHKEREAAELGELISIEK--GFVGPYHSHKEMPHLVNLSDDPLLAEL 509
DB 455 netweeklkrtealrmereallaemgavvredgdtgvgfppkktphlvnlndplmsecl 514
QY 510 VYNIKPGQTRGVNVDQAEIRLNGSKILKEHCTPENV-----DNVVTIVPNEKAAVMV 564
DB 515 lyhikdgtrvgvgvmdm----ikltgqfirqehclfrsibpqpdegvvtlepecegaetyv 570
QY 565 NGVRIKPTLRSGYRIILGDFHFRNPEEAREARQEQSLRHVSNSQLGSPAPGRH 624
DB 571 ngklvtleplvksnrlvmgmhvfrrfnheqarlerer-----gvppp--- 614
QY 625 DRTLKAGSDAGDSRSDSLPHFRGKDSDFWYARREAAASAILGLDOKISHLTDDELDA 684
DB 615 -----pgppsepvdwnfaqllek-qgidikle-mekrlqdl 650
QY 685 FDDVQKARAVRRGLVEDNE---DSDS 707
DB 651 enqyrkekeeadllegqrllyadsds 676
RESULT 5
AAE04316
ID AAE04316 standard; Protein; 1103 AA.
XX
XX AAE04316;
AC

XX WIPI; 2001-656860/75.
DR N-PSDB; ABL08011.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 18516; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1773 AA;

Query Match 40.6%; Score 1635; DB 22; Length 1773;
Best Local Similarity 44.5%; Pred. NO. 1.2e-115;
Matches 367; Conservative 143; Mismatches 193; Indels 122; Gaps 20;

QY 6 NIKVVVRFPNAREIDRGAKCIVRMVGNQITLTPPGAEEKARKSGKTKMDGPKAFAD 65
DB 32 svkvavrvrpfnsrelarekskleimagattatnp-----kvpntsdsvkrfnfd 83

QY 66 RSYWFDKNAPYARQBDLQDLGVLPLLDNAFGYNCFAYGQTGSGKSYMMGYGKE- 124
DB 84 ysywshdhddadstgamykylgeemlqbsfgdygncvifaygqtgagkysymgrqeeq 143

QY 125 -HGVIPIRCODMFRINELQKKNLCTVEVSYLEIYNERVOLLNPSTKGNLKVREHPS 183
DB 144 qeglipmickdltrigdtdt-dlkysvevsymelycervrdlnpknkgnlrvehpl 202

QY 184 TGPVEDLAKLVRSFOEINLMDENKARTVAATNNETSSRSRAVFTTLTKQWDEE 243
DB 203 lgpvyedsklavdygdihlddegknartvaatmmetsrshavfttffqrrhdm 262

QY 244 TKMDTEKAVISLVDLAGSRATSTGATGARLKEGAEINSLTGLRVIAALADMSSGKQ 303
DB 263 tnlttekvsislvdlagseradstgagtrlkeganinksltlglkvisalaevasskk 322

QY 304 --KNQLVPYRDSVLTLWLLKDSLGGNSWTAMIAISPADINFEETLSTLYADSARKIKN 361
DB 323 ntkkadfipyrdtsaltwllrenlggnsktamaiaispadinydetlstlryadrakqvc 382

QY 362 HAVVNEDPNARMIRELKEELQAULRSKLOSGGGGAGGSGGPVEESYPDPDTLEKQIVS 421
DB 383 kavnedanaklrelkeelqirdllkaeg-----levqegpdgk-----vv 425

QY 422 IQQPDATVKKMSKAEI-----VEOLQSEKILYRDLNQTWEKLAKE 463
DB 426 cekdrankdelstkvtsptksnrngsttemavdqkasekliaelnetweeklkrte 485

QY 464 EHKEREAALEELGISI-EKGF-VGPHSKEMPHLVNLSDDPLLAELVYNIRPGQTRVG 521
DB 486 eirvgvaeaemgavkedgltvgvfpktpplhvalnedpnlsecllyyikegltrlg 545

QY 522 NVNODQAEIRLNGSKILKEHCTFENVDNVTIVPNEKAAVMVNGVRIDRTPRLRSYRI 581
DB 546 theanvpqdqlqslgshlkehtctfenkstvtllphkdaiiyngrklvepevrlktgrv 605

QY 582 ILGFHIFRPNHPEAREAEQESLLRHSVTNSQLGSPACGRHDRTLSKAGSDADGDSRS 641
DB 606 ilgknhvfrfnpegare-----lrdki-----eteneaeeneve---kt 641

QY 642 DSPLPHFRGKDSWVFYARREASAILGLD-----OKISHLTDD-----ELDALFDVQ 689
DB 642 dt-----gqvdmfagcellek-qgidlkaemkkrlndleeqykreklaqdgqfeeqr 693

QY 690 K---AR--AVRRGLVEDNEDSDSQSPFVRDVKYMSNGTIDNFSLDYTAITMP----- 735
DB 694 ktyearidalqkveeqetmsmysyspedfhqeedytnpnyescwtareaglaawaf 753

QY 736 -----GTPSRDDGDDALFFGDKSKQDASNVVDVEELRQQAQ 772
DB 754 rkwyrbqftslrddlwgnaifl-----keanaaisvelkkkvqfq 792

RESULT 7
AAU19569
ID AAU19569 standard; Protein; 757 AA.
XX
XX AAU19569;
XX
XX 04-DEC-2001 (first entry)
XX
XX Human diagnostic and therapeutic polypeptide (DITHP) #155.
XX
XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
XX cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
XX acquired immune deficiency syndrome; AIDS; autoimmune disorder;
XX respiratory disorder.
XX
XX Homo sapiens.
XX
XX WO200162927-A2.
XX
XX 30-AUG-2001.
XX
XX 21-FEB-2001; 2001WO-US06059.
XX
XX 24-FEB-2000; 2000US-0184693.
XX 24-FEB-2000; 2000US-0184897.
XX 24-FEB-2000; 2000US-0184898.
XX 24-FEB-2000; 2000US-0184768.
XX 24-FEB-2000; 2000US-0184769.
XX 24-FEB-2000; 2000US-0184770.
XX 24-FEB-2000; 2000US-0184771.
XX 24-FEB-2000; 2000US-0184772.
XX 24-FEB-2000; 2000US-0184773.
XX 24-FEB-2000; 2000US-0184774.
XX 24-FEB-2000; 2000US-0184776.
XX 24-FEB-2000; 2000US-0184777.
XX 24-FEB-2000; 2000US-0184797.
XX 24-FEB-2000; 2000US-0184813.
XX 24-FEB-2000; 2000US-0184837.
XX 24-FEB-2000; 2000US-0184841.
XX 24-FEB-2000; 2000US-0185213.
XX 24-FEB-2000; 2000US-0185216.
XX 12-MAY-2000; 2000US-0203785.
XX 15-MAY-2000; 2000US-0204226.
XX 16-MAY-2000; 2000US-0204525.
XX 16-MAY-2000; 2000US-0204821.
XX 16-MAY-2000; 2000US-0204908.
XX 16-MAY-2000; 2000US-0205232.
XX 17-MAY-2000; 2000US-0204815.
XX 17-MAY-2000; 2000US-0204863.
XX 17-MAY-2000; 2000US-0205221.
XX 17-MAY-2000; 2000US-0205285.
XX 17-MAY-2000; 2000US-0205286.
XX 17-MAY-2000; 2000US-0205287.
XX 17-MAY-2000; 2000US-0205323.
XX 17-MAY-2000; 2000US-0205324.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Panzer SR, Spiro PA, Barville SC, Shah P, Chalup MS, Chang SC;
PI Chen A, D'Sa SA, Amshey S, Dahl CR, Dam TC, Daniels SE;

PN	WO200061629-A1.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				</
----	-----------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

Qy 357 KRIKHAVVNDPNRMIREKEELAKLRSLQSSGGGGAGGGGPPVEESYPDPPTPLE 416
Dy 307 krivhavvndpnrmirekeelaklrslqssggggagggppveesypdpptple 416
Qy 417 KQIVSIQOPDATVKKMSKAEIVQNLQSEKLYRDLNQTWEEKLAKTEIHKEREAALEEL 476
Dy 347 -----gdvqdklaesenlmkqlsgtweeklvktrigerngqalekm 388
Qy 477 GISIEKGFVGPYHSHKEMPHLVNLSDDPLLAELCLVYNIKPGQTRVGNVNDQQAERLNGS 536
Dy 389 gisvqas--gikvcknkylylnadpslnellvyyikdrtligrtsigqpdqisgl 446
Qy 537 KILKEHCTFENVNVVTVIPNEKAAMVWVGRIDKPTLRSLGYRIILGDFHFRFNHP 594
Dy 447 gikpehcvitiedsglymepvggarcfvgaaavektpmqgdrllwgnhhrfvnsp 504

RESULT 10
AAB63190
ID AAB63190 standard; Protein: 503 AA.
AC AAB63190;
XT 26-MAR-2001 (first entry)
XX Human secreted protein sequence encoded by gene 5 SEQ ID NO:116.
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disease; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
KW food additive; preservative.
XX Homo sapiens.
XX W0200061629-A1.
XX 19-OCT-2000.
XX 06-APR-2000; 2000WO-US09071.
XX 09-APR-1999; 99US-0128694.
XX 20-JAN-2000; 2000US-0176931.
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX Ruben SM, Komatsoulis G;
XX WPI; 2000-647420/62.
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX Disclosure; Page 489-491; 533pp; English.

CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular
CC disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral
CC ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's
CC disease, infections caused by bacteria, viruses and fungi and ocular
CC disorders e.g. corneal infection. The polypeptides can also be used to
CC aid wound healing and epithelial cell proliferation, to prevent skin
CC aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. AAF22364 to
CC AAF22372 and AAB63133 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 503 AA;
Query Match 31.7%; Score 1276; DB 21; Length 503;
Best Local Similarity 50.8%; Pred. No. 5.3e-89;
Matches 277; Conservative 74; Mismatches 136; Indels 58; Gaps 11;
Qy 58 GP-KAFADFYSYFDKNA-PNYARQEDLFODLGVPLLDNAFKGYNNCIFAYGQTGSGKS 115
Dy 9 gpaevfaydhcfwsmdesvkekygqdivfclgenilqnaifgynacifaygqtcgsks 68
Qy 116 YSMGKGKEHGVIPIRICODMFERRI----NELQKDKNLCTVEVSYLEIYNRVRDLLNP- 170
Dy 69 ytmngtdqpgliprlcslgflertqkegneesgfk-----vevsymelynekvrdllqk 123
Qy 171 STGNLKVREHPSGPGYVEDLAKLVRSFOEINLMDGNCARTVAATNMETSRSRAV 230
Dy 124 gsrqtklvrehsvigpyvgdlskavtsykdiesslmsegnskrtvaatnmneessrshav 183
Qy 231 FTLTLTQKWHDEETKMDTEKVAKISLVDLGASERATSTGATGARLKEAGAEINRSLSTLGR 290
Dy 184 fkitlthlyvsksgtsgskgkxslvdxgeratkgaagdrilkegnslnksltlgl 243
Qy 291 VIALADMSGKQKKNQLVPYRDSVLTWLLKDSLGNSMTAMIAAISPADINFEETLSTL 350
Dy 244 visaladqsagk--sxnkfpyrdsvlwtllkdslgnskxamvatvspaadnydetlsl 302
Qy 351 RYADSARKIKHAVVNDPNRMIREKEELAKLRSLQSSGGGGAGGGGPPVEESYP 410
Dy 303 ryadrakhivhavvndpnrmirekeelaklrslqssggggagggppveesyp 410
Qy 411 PDTPLEKQIVSIQOPDATVKKMSKAEIVQNLQSEKLYRDLNQTWEEKLAKTEIHKERE 470
Dy 344 -----eamkspeldrieeseekligemtvtweeklrteelaqerq 384
Qy 471 AAELGISTIEKGFVGPYHSHKEMPHLVNLSDDPLLAELCLVYNIKPGQTRVGNVNDQQA 530
Dy 385 kqleslgisqlqss--gikvgddkcfvlnadpslnellvyyike-htligsansq---d 438
Qy 531 IRLNGSKILKEHCTFE-NVDNVVTVIPNEKAAMVWVGRIDKPTLRSLGYRIILGDFHIF 589
Dy 439 iqicgmglilbehcliditseggmtpqkntrtfvngssvspqilghgdrilwgnhff 498
Qy 590 RFNHP 594
Dy 499 rlnlp 503
RESULT 11
AAM41820
ID AAM41820 standard; Protein: 421 AA.
XX AAM41820;
XX 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 6751.
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX Homo sapiens.
OS
XX WO200153312-A1.
PN
XX 26-JUL-2001.
PD
XX 26-DEC-2000; 2000WO-US34263.
PF
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-052317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang JT, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou F, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR N-PSDB; AAI60976.
DR
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 6751; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAI38642-AAI42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
SQ Sequence 421 AA;

Query Match 31.0%; Score 1251; DB 22; Length 421;
Best Local Similarity 61.1%; Pred. No. 3.3e-87;
Matches 245; Conservative 68; Mismatches 64; Indels 24; Gaps 6;

QY 4 GGNIKVVVRFPNAREIDRGAKCIVRMENQOTILTTPPGAEERKARKSGKTMIDGPKAFA 63
DB 9 gasvkvavrvpfnsremrsdskciqmsgstttivnpkpket-----pkfsfs 57
QY 64 FDRSWSFDKNAP---NYARQEDLFODLVPLLDNAFKYNNCIFAYGOTGSGKSYSNMG 120
DB 58 fdysyws--htspedinyasqkvqyrdigeemlqnafegynvcifayqyqtagksytumg 115
QY 121 YGK--EHGVIPRICDMFRRNELQKKNLACTVEVSYLEYNERVRDLLNPSTKGNLKV 178
DB 116 kqeqdggqglipqlcdlfsrindttnd-nmsysvsvsmeiycervrldllnpknknlrv 174

QY 179 REHSTGTPYVEDLAKLVVRSEFOETIENLMDGKNKARTVAATNNNETSSSHAVFTLTQK 238
DB 175 rehpllgpyvedsklavtsyndiqldmdsgnkartvaatnmnetssrshavnifftqk 234
QY 239 WHDETKMDTEKVAKISLVDLAGSERATSTGATGARLKEGAEINRSLSTLGRVIAALADM 298
DB 235 rhdaetnlttekvsksislvdiagseradstgagtrlkeganinkslctllgkvisalaem 294
QY 299 SSG-----KQKNQLVYPYRDSVLTWLLKDSLGNSMTAMIAAISPADINFEETLSTLYA 353
DB 295 dsppnknkkkktfdpyrdsvtllrenlgnnartamvaalspadinynetlslrya 354
QY 354 DSAKRIKNHAVNEDPNARMIRELKEELAQLRSKLQSSGGG 394
DB 355 draklrcnavinedpnknklrelkdevtrldliyaaglg 395

RESULT 12
ABB61704
ID ABB61704 standard; Protein; 1174 AA.
XX
AC ABB61704;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 11904.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL05807.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
PS Disclosure; SEQ ID NO 11904; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1174 AA;

Query Match 28.0%; Score 1128; DB 22; Length 1174;
Best Local Similarity 36.5%; Pred. No. 4.1e-77;
Matches 286; Conservative 125; Mismatches 267; Indels 106; Gaps 21;

Qy 25 ACIVRMGEGNOTILTPPGAEKARKSKTMDGPKAFAPDRSYWSFDKNAPNYARQBDL 84
Dd 3 aqlimemenktrllkp-riqsirdegr---dnhhdfdyyswfsdaedphfatgeqv 57
Qy 85 FODLGVPLLDNFAKGYNNCIFYAGTGGKSYMMGYGKEHGVIPRICQDMFRRLNELQK 144
Dd 58 ysdlgndvvdcaeyegynacvfayggtgsgkftmgtgtpnpglpriceelfarm-rvvg 116
Qy 145 DKNLTCTVEVSYLEYINRVRDLNPSYKGN-LKVREHPTGPGYVEDLAKLVRSFQIE 203
Dd 117 esgtgyrthasylelynervkdliaaagstghglrvrehslrgpyvenlsqhvafdeiq 176
Qy 204 NLMDEGNKARTVAATNNMETSRSRHAVFTLTLTKQHHDEETKMDTEKVAKISLVDLAGSE 263
Dd 177 eciargnqrftastnmndtsrshafitfvqgvfmd--mpsetvskhlvdlagse 234
Qy 264 RATSTGATGARLKEGAENRSLSTLGRVIAALADMS-SGOKKKNOLVPRDSVLTWLKLD 322
Dd 235 ranatgatgqrllkegahinkslvtlgsvisalaeqggghssssyipyrdsliltwlkld 294
Qy 323 SLGNSMTAMIAAISPADINFEETLSTLYADSARKIKNHAVVNEDPNARMIRELKEELA 382
Dd 295 slggnsktimiaaspacnyssetlstryanrakniinkptvnedtnvklireeiein 354
Qy 383 QLRSKLQSSGGGGGAGCGGPGVEESYPPDPLEKQIVSIQPDATVKKMSKAEIVEQLN 442
Dd 355 klskml-----agdihsllqp-----slkviadlq 378
Qy 443 QSEKLYRDLNOTWEKLAKEIHKREAALEELGISIEKGFVGPYHSEKMPHLVNLSD 502
Dd 379 kkeaqekvlteewekwkvagilqeks-----lgirkgvgvvlsemphlighnd 432
Qy 503 PLLAELVYNTKPGTRGVNQNQDTQAEIRLNGSKILKEHCTFENVNVYTVIPNEKAAV 562
Dd 433 -vttgtvlylsikegetrigsedadvagdielagdgiragchcsiflkggvvtlhpwplaqc 491
Qy 563 MVNGVRIDKPVRLSGYRIILGDFHIFRNHPPEARERQESLLRSHSVNSQLGSPAG 622
Dd 492 wvnhlldepqieggdillig-----qdlsrqlmsrlslits 532
Qy 623 RHDRTLKAGSDADGSDSDSPLPHFRGKDSDFYARREA---SAILGLQOKISHLDD 679
Dd 533 kenlitsiysdedgas-----pykrperq-yyprqmsrddpelqdenrkildtien 584
Qy 680 ELDAL-FDDVQ-----KARAVRGLVE-DNEDSDSSSPVRD-----KYMNGITD 724
Dd 585 alkqinvervqmhdyktkvkrkteelirleqeemdgqlincreeqliarkdmllwekn 644
Qy 725 NFSLDTATM-PGTPRSDD-----DGDALFGD-KKSKQDASNVVDVELRQOQQAQMEEA 776
Dd 645 nekvqgmkmieegtbindellllqvdsdldlfaaqfiktvtvrrnnneirkldeqlaekerl 704
Qy 777 LKTA 780
Dd 705 lnas 708
RESULT 13
ID ABB72021
XX AC ABB72021 standard; Protein; 1121 AA.
XX AC ABB72021;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 42855.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
OS Drosophila melanogaster.
XX W0200171042-A2.

XX 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US09231.
PF 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PI Venter JC, Adams M, LI PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL16124.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure; SEQ ID NO 42855; 2lpp + Sequence Listing; English.
PS The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1121 AA;
Query Match 25.7%; Score 1035; DB 22; Length 1121;
Best Local Similarity 33.7%; Pred. No. 5e-70;
Matches 275; Conservative 131; Mismatches 293; Indels 116; Gaps 21;
Qy 1 MSGGGNIKVVVRPFNFNAREIDRG-AKCIVRMEGNOTILTPPGAEKARKSGKTMDGP 59
Dd 115 vseeemlvavvrplnalectrigqvtvqvghnsneltvqagssada-sagvt----- 168
Qy 60 KAFADRYSWSPDKNAPYARQEDLFQDLVPLLDNAFKNYCNCFAYGQTGSGKSYMM 119
Dd 169 hffsydyvyyscdperknfacqakvfegtarplidatfegynacifayqgtgsgksymm 228
Qy 120 GY-----GKEH---GVIPRICQDMFRRLNELQDKNLTCTVEVSYLEYINRVRD 166
Dd 229 gtealddaaldggpphdeagilprfchelfrrieavksqqqlqveveysfyfelyneklnd 288
Qy 167 LLN-----PSTKGNLKVREHPTGPGYVEDLAKLVRSFQIEINLMD 208
Dd 289 llsvqhaaaatgestpiqqqqqqqpalkvrehpifgpyvvdlsahsvdsysalnrlwlv 348
Qy 209 GNKARTVAATNNMETSRSRHAVF--TLTLTKQHHDEETKMDTE-----KVAKISL 256
Dd 349 gnsqratastamndksrshsifnvlntldssddglsdtdssatasslrqtrrsksl 408
Qy 257 VDLAGSERATSGATGARLKEGAENRSLSTLGRVIAALADMS-----GKQKKNQLV 309
Dd 409 vdlagserisvsgngiregvsinkslitlgkvaaladskakalangplsgtstfv 468
Qy 310 PYRDSVLTWLKLDLSLGGNSMTAMIAAISPADINFEETLSTLYADSARKIKNHAVVNEDP 369
Dd 469 pyresvltwlrrenlggnsktvmatlaspashadetiattrackarsivnrkvnesp 528
Qy 370 NARMIRELKEELAQLRSKLQSSGGGGGAGGSGGPGVEESYPPDPLEKQIVSIQPDATV 429
Dd 529 hdkliirdraevdrikslrneyerqrrisgnsnnpvrkiliet-----svdetevea 581
Qy 430 KMSKAEIVEQLNQSEKLYRDLNQTWEKLAKEIHKREAALEELGISIEKGFVGPYH 489

```

Db 582 lrqlaerereisraqk-----smexlkeadqrkselrvlkrqglalel-----ta 629
QY 490 SKEMPHLVNLDPLAEGVLVNIKPGQTRVG-----NVNQDTQAEIRLNGSKILKEHCT 544
Db 630 eqkacqlvntadpilsqgtfylvlpqglrvlrgrrlpqsgsssqdpdivldgplvalghcs 689
QY 545 FENV--DNVVTIVPNEKAAVMVNGVRIDKPTRLRSGYRIILGDHFIFRNP--EARA 599
Db 690 ieherggklyvpqgsedfetyvngellkrrqlfhgdrivigshyfrlspnfcsggrka 749
QY 600 EROEQSLLRH-----SVTNSQLGSPAPGRHRTLSKAGSDADGSDSDS-----PLPHPR 649
Db 750 dhpvdqlahqellkqegqlrseleaeakraaltkieqeraghardfeerlqcleleqfk 809
QY 650 GK--DSOWFYARRAASAILLDQKISHLTDELDALFDDVQKA--RAVRRLGIVEDNDS 705
Db 810 ykcnsemleterda---lalaqqghehtplrhedavstpaqkstiliedlqrimlpsees 865
QY 706 DSQSFVRD-----KYSNGTIDNFSL 728
Db 866 lhktqlmvkeatqrcqldlplefrtqtqdpdefgl 900

RESULT 14
AAW03691
ID AAW03691 standard; Protein; 955 AA.
XX
AC AAW03691;
XX
DT 08-MAR-1995 (first entry)
XX
DE K39 polypeptide of Leishmania chagasi.
XX
KW Parasite; leishmaniasis; infection; repeat unit; vaccine; antigen;
KW Leishmania chagasi.
XX
OS Leishmania chagasi.
XX
PN WO9416331-A.
XX
PD 21-JUL-1994.
XX
PF 10-JAN-1994; 94WO-US00324.
XX
PR 15-JAN-1993; 93US-0006676.
XX
PA (IASY-) IASYS CORP.
XX
FI Read SG;
XX
DR WPI; 1994-249402/30.
DR N-PSDB; AAQ70152.
XX
PT Diagnosis of Leishmaniasis - by determining the presence of
PT antibodies that bind to a K39 repeat unit antigen
XX
PS Disclosure; Page 12-15; 28pp; English.
XX
CC The K39 polypeptide comprises a number of repeated units (described
CC in AA057366). Detection of antibodies directed against this repeated
CC unit in a patients sample is indicative of leishmaniasis. The
CC antigenic repeat unit can itself be used as a vaccine to protect
CC against infection by a leishmania parasite.
XX
SQ Sequence 955 AA;

Query Match 20.2%; Score 814; DB 15; Length 955;
Best Local Similarity 28.1%; Pred. No. 3.2e-53;
Matches 253; Conservative 125; Mismatches 292; Indels 230; Gaps 24;

QY 7 IKVVVRVRFNARE--IDRGAKCIVRMEGNQITLTPP-----PGAEKARKSGKTMD 57

```

```

Db 13 VKSVVRVPLNERENNAPGCTKVTVAQKAAAVVTKVVGSSNNNSGAESMGTRARRVAQD 72
QY 58 GPKAFAPDRSYWSF---DKNAPNYARQEDLFQDLGVPLLDNAPKGYNNCIFAYCQOTSGK 114
Db 73 ---fqdhvfwsvetpdacatpatqadvrtigpylvqhafgdfnsclfayqgtsgk 128
QY 115 SYSMMG-----YKHEGVIPRICODMPRRINELOKKNLCTCTVEVSYLEIYNERVRL 168
Db 129 tytmmgadvsalsgegnvtptricleifarkeasvgaqhsrwivelgyveynervsdl 188
QY 169 NPSTGK-----NLKYREHPSTGPIYVEDLAKLVVRSEFOEINLMDGNKARTVAATNM 220
Db 189 gkrkkgvkggeevydvrenpsrfgvlegrlvevgsladdvvrllieignvritastkm 248
QY 221 NPTSSRSHAVFTLTL---TQKHDEETKMDTEKVAKISLVLDLAGSERATSTGATGARLK 276
Db 249 ndrsrshaimlllreertmttkagetirtagkasrnmnlvdlagservaqsgvqgfk 308
QY 277 EGAEINRSLSTLGRVIAALADMS--GKOKKNOLVYPRDSVLTWLLKSLGGNSMTAMTAA 335
Db 309 eathlnslttlgrvldmadmatkgakqsvapfrskitfllkdslgnsktfmat 368
QY 336 ISPADINPEETLSTLYADSAKRIKNHVVNEDPNARIRELKEELAQLRSKLSQSGGG 395
Db 369 vspalnlyeetlstlryasaradivnvagvnedprarrileeqmedmrqam----- 421
QY 396 GGAGSGGGPVVE-----SYPTPTPLEK 417
Db 422 --agdpayvselkklalleseaqkraadlqalererehngvqerllrateakeles 479
QY 418 QIVSTQ-----PDATVKMKSAEIVEQL 441
Db 480 raalqeemtatrqqdkmqalnrlkeeqarkerellkemakkaalskvrirkaeia 539
QY 442 NOSEKLYDLNQTWEKLAKEEIIHKEREAALEELGISTIEKGFVGPYHSKEMPHLVNLS 501
Db 540 serek1-----estvaqlereqrevaldal-----qthqrkiquealesse 581
QY 502 -----DPLLAELVYNIKPGQTRVGNVNDTQAEIRLNGSKILKEHCTFENVVNVTV 555
Db 582 rtaaedqllq--ltelqsertqlsqvvtredrltr-dlqriqveygetelardvalca 638
QY 556 PNE-----KAAVMVNGVRIDKPTRLRSGYRIILGDHFIFRNPHEARAEQSLLRHS 610
Db 639 agemearyaavfhlqtllelatawedair-----eralaeadeaaaae-- 682
QY 611 VTNSQLGSPAPGRHRTLSKAGSDADGSDSDSPLPHFRGKSDWIFYARREASAILGLD 670
Db 683 -----ldaaastsqu-----aresaceritsle 705
QY 671 QKISHLTDELDALFDDVQKARAVRRLGIVEDNEDSDSQSFVPVRKYMNGTIDNFSLDT 730
Db 706 qqlre-seeraaelasqleataaaksaqcdrentatleqqires-----earaael 757
QY 731 AITMPTPSRDDDDGDLFFGDKKSKQDASNV-----DVEELRQQAQMEALKTKAR 781
Db 758 asqleataaa-----kmsaqcdrentatleqqirdseeraaelasqlestaak 807

RESULT 15
AAW03691
ID AAW03691 standard; Protein; 955 AA.
XX
AC AAW03691;
XX
DT 09-MAR-1997 (first entry)
XX
DE Leishmania chagasi K39 antigen.
XX
KW Leishmania chagasi; acidic ribosomal antigen; Lcp0;
KW epitope; K39.
XX

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2002, 04:39:26 ; Search time 25.38 Seconds
(without alignments)
754.518 Million cell updates/sec

Title: us-09-235-416-1

Perfect score: 4030

Sequence: 1 MSGGNIKVVVRPFNARE.....ELRQQQAQMEALKAKQEF 784

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1658	41.1	1103	3	US-09-162-373-1
2	1658	41.1	1103	4	US-09-467-946-1
3	1311.5	32.5	1375	4	US-09-722-139-2
4	1005.5	25.0	563	4	US-09-718-841-2
5	991	24.6	359	4	US-09-722-139-4
6	814	20.2	955	1	US-08-006-676B-1
7	814	20.2	955	1	US-08-282-845-2
8	814	20.2	955	5	PCT-US94-00324-1
9	812.5	20.2	955	2	US-08-428-414A-3
10	747	18.5	357	4	US-09-718-841-4
11	673	16.7	1057	4	US-09-541-782-10
12	664.5	16.5	1388	4	US-09-572-191-2
13	640.5	15.9	409	4	US-09-572-191-6
14	633	15.7	1184	4	US-09-541-782-2
15	625	15.5	411	2	US-08-713-815A-4
16	625	15.5	441	2	US-08-713-815A-3
17	610.5	15.1	1066	4	US-09-541-782-8
18	591	14.7	375	4	US-09-572-191-4
19	587.5	14.6	1038	4	US-09-541-782-4
20	579	14.4	1073	4	US-09-541-782-6
21	541	13.4	299	4	US-09-621-233-2
22	513.5	12.7	790	4	US-09-723-153-2
23	464	11.5	326	4	US-09-721-383-4
24	464	11.5	787	4	US-09-721-383-2
25	457.5	11.4	338	4	US-09-723-153-4
26	304.5	7.6	1201	4	US-09-098-901-2
27	258	6.4	205	4	US-09-637-481-2

28	195	4.8	1612	3	US-08-545-860D-48	Sequence 48, Appl
29	195	4.8	1612	5	PCT-US94-04496-48	Sequence 48, Appl
30	185.5	4.6	1829	4	US-09-157-420-1	Sequence 1, Appl
31	143	3.5	3248	1	US-08-353-700-1	Sequence 1, Appl
32	143	3.5	3248	5	PCT-US95-16216-1	Sequence 1, Appl
33	142	3.5	2482	1	US-08-328-254-6	Sequence 6, Appl
34	140	3.5	1093	5	PCT-US93-03077-1	Sequence 1, Appl
35	138	3.4	1354	3	US-08-685-871-2	Sequence 2, Appl
36	137	3.4	976	4	US-09-104-324B-4	Sequence 4, Appl
37	136	3.4	1388	2	US-08-685-576-4	Sequence 4, Appl
38	135.5	3.4	700	1	US-07-720-589-2	Sequence 2, Appl
39	135.5	3.4	700	5	US-08-785-190-2	Sequence 2, Appl
40	135.5	3.4	700	5	PCT-US92-05539-2	Sequence 2, Appl
41	134.5	3.3	859	4	US-09-199-637A-281	Sequence 281, App
42	133.5	3.3	700	4	US-08-235-836C-66	Sequence 66, Appl
43	133	3.3	1388	2	US-08-685-576-1	Sequence 1, Appl
44	132.5	3.3	1702	4	US-08-296-791-5	Sequence 5, Appl
45	132.5	3.3	1702	5	PCT-US95-10661A-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-162-373-1
; Sequence 1, Application US/09162373
; Patent No. 6013454
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: KINESIN-LIKE MOTOR PROTEIN
; FILE REFERENCE: PF-0593 US
; CURRENT APPLICATION NUMBER: US/09/162,373
; CURRENT FILING DATE: 1998-09-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 1103
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 1281811
US-09-162-373-1

Query Match	41.1%	Score 1658;	DB 3: Length 1103;
Best Local Similarity	47.3%	Pred. No. 5.1e-131;	
Matches	353;	Conservative 120;	Mismatches 159; Indels 114; Gaps 16;
Qy	4	GNIKVVVRPFNAREIDRGAKCIVRMENGTILTPPPGAEEKARKSKGKTMDGPKAFA	63
Db	3	GASVQVAVRPFNAREISQDAKCVVSMQNTSINP-----KQSKDAKPKSET	51
Qy	64	FDRSTWSFDKNA-PNYAKQEDLFQDLGVPLDPAKGVNVCIFAYGQTSGSKSYMMGVG	122
Db	52	FDSYSSWSTSTEDPOFASQOQVYRDIQGEMLLHAFEGYVNCIFAYGQTGAGKSYTMGRQ	111
Qy	123	K--EHGVIPRICODMFRINELQDKNLCTCTVEVSLEYINERNVRLDLPSTKGNLKYRE	180
Db	112	EPGQGVIPQCEDEDFSRVSENG-SAQLSYSVEVSMEIYICERVRLDLPKRGSLRVRE	170
Qy	181	HPSTGPYVEDLAKLVRSFOEINLMDGNKARTVAATNMNETSSRSNAVFTLTITQKWH	240
Db	171	HPILGPYQDLSKLVTSYADIADLMDCGNKARTVAATNMNETSSRSNAVFTIVTQRCH	230
Qy	241	DEETKMDTEKVAKISLVLDLAGSERATSTGATGARLKEGAEINRSLSLTGRVTAALADMSS	300
Db	231	DQLTGLDSEKVKISLVLDLAGSERADSSGARGMGLKEGANINKSLTTLGKVISALADMQS	290
Qy	301	GQKKNQLVPPYRDSVLTWLLKDSLGNSMTAMIAISPADINFETLTSLRVTADSARKIK	360

Db 291 -KKRSDFIPYRDSVLTWLLKENLGNSTRTAMIAALSPADINYEETLSTLYADRTKQIR 349
QY 361 NHAVNEDPNARMIRELKEELAQSLKSSGGGGGAGG----- 400
Db 350 CNAIINEDPNARLIRELQEVARRELLMAOGLSASALEGLTEGSGVGPALPVSPPA 409
QY 401 -----SGGPVSESYPPDTPLEKQIVSIQQPDATVKMKAIEVBNQSEKLYRDL 451
Db 410 PVSPPSTTHNGELEPSPNT--ESQI-----GPEAMERLQETEKIIAEL 454
QY 452 NOTWEEKLAKTEEIHKEERAALEELGISIEK--GFVGPYHSEKEMPHLVNLSDDPILAECL 509
Db 455 NETWEEKLRKTEALRMEREAALLAEMGVAVREDGGTVGVFSPKKTPLHLNLEDPLMSECL 514
QY 510 VYNIKPGQTRVGNVODTQAEIRLNGSKILKEHCTFENV-----DNVVTIVNEKAAVMV 564
Db 515 LYHIKGVTRVGOVDM-----IKLTQFIREQHCLFRSIPQDPGEVVVTLPECEGAETV 570
QY 565 NGVRIDKPTRLRSYRIILGDFHIFRNFHPEARAEQSLRLHRSVTSNLSQSGSPAPGRH 624
Db 571 NGKLVTPLVLKSGNRIVMGKNHVRFNHPEQARLER-----GVPPP----- 614
QY 625 DRTLKAGSDADGSDSDPLPHFRCKSDWYARREASAILGLDOKISHLTDDDELDAI 684
Db 615 -----PGPPEPVDWNPFAQKELLEQ--QGIDIKLE--MEKRLQDL 650
QY 685 FDDVQKARAVRRLGVEDNE---DSDS 707
Db 651 ENQYRKEKEADLLEQORLYADSDS 676

RESULT 2
US-09-467-946-1
; Sequence 1, Application US/09467946
; Patent No. 6248594
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: KINESIN-LIKE MOTOR PROTEIN
; FILE REFERENCE: PF-0593 US
; CURRENT APPLICATION NUMBER: US/09/467,946
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: 09/162,373
; EARLIER FILING DATE: 1998-09-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 1103
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1281811
US-09-467-946-1

Query Match 41.1%; Score 1658; DB 4; Length 1103;
Best Local Similarity 47.3%; Pred. No. 5.1e-131;
Matches 353; Conservative 120; Mismatches 159; Indels 114; Gaps 16;

QY 4 CGNIKVVVRPFNFNAREIDRGAKEIVRMENQITLTTPPGAEEKARKSGKTMIDGPKAFA 63
Db 3 GASVAVVVRPFNFNARETSQDAKCVMSQGNNTSIINP-----KOSKDAKSGFT 51
QY 64 FDRSYWSFDKNA-PNYARQEDLFQDLGVPLLDNAFKYNNCIFAYGTGSGKSYSMGYG 122
Db 52 FDIYSWSTSTEDPQFAQQQYVRDIEGEMLLHAFEGYNVCIFAYGTGAGKSYTMGRQ 111
QY 123 K--EHGVIPRICODMFRNELQDKNLTCTVSVSYLBIYNERVRLNLPSTKGNLAVRE 180
Db 112 EPGQQGIVPOLCEDLFSRVSENO--SAQLSYSEVSMEIYECRVRDLNLPKSRGSLRVRE 170

QY 191 HPSTGPPYVEDLAKLWRSFQRIENLMDGKNKARTVAATNMNETSSRSHAVFTLLTOKWH 240
Db 171 HPILPGYQDLSKLAFTSYADIADLMDCGNKARTVAATNMNETSSRSHAVFTIYFTQCH 230
QY 241 DEETKMDTEKVAKISLVDLAGSERATSTGATCARLKEGAENRSLSTLGRVIAALADMS 300
Db 231 DOLTGLDSEKYSKISLVDLAGSERADSSGARGMGLKEGANINKSLTTLGKVISALADMS 290
QY 301 GKQKNQLVPYRDSVLTWLLKDSLGGNSMTAMIAAISPADINFEETLSTLYADRTKQIR 360
Db 291 -KKRSDFIPYRDSVLTWLLKENLGNSTRTAMIAALSPADINYEETLSTLYADRTKQIR 349
QY 361 NHAVNEDPNARMIRELKEELAQSLKSSGGGGGAGG----- 400
Db 350 CNAIINEDPNARLIRELQEVARRELLMAOGLSASALEGLTEGSGVGPALPVSPPA 409
QY 401 -----SGGPVSESYPPDTPLEKQIVSIQQPDATVKMKAIEVBNQSEKLYRDL 451
Db 410 PVSPPSTTHNGELEPSPNT--ESQI-----GPEAMERLQETEKIIAEL 454
QY 452 NOTWEEKLAKTEEIHKEERAALEELGISIEK--GFVGPYHSEKEMPHLVNLSDDPILAECL 509
Db 455 NETWEEKLRKTEALRMEREAALLAEMGVAVREDGGTVGVFSPKKTPLHLNLEDPLMSECL 514
QY 510 VYNIKPGQTRVGNVODTQAEIRLNGSKILKEHCTFENV-----DNVVTIVNEKAAVMV 564
Db 515 LYHIKGVTRVGOVDM-----IKLTQFIREQHCLFRSIPQDPGEVVVTLPECEGAETV 570
QY 565 NGVRIDKPTRLRSYRIILGDFHIFRNFHPEARAEQSLRLHRSVTSNLSQSGSPAPGRH 624
Db 571 NGKLVTPLVLKSGNRIVMGKNHVRFNHPEQARLER-----GVPPP----- 614
QY 625 DRTLKAGSDADGSDSDPLPHFRCKSDWYARREASAILGLDOKISHLTDDDELDAI 684
Db 615 -----PGPPEPVDWNPFAQKELLEQ--QGIDIKLE--MEKRLQDL 650
QY 685 FDDVQKARAVRRLGVEDNE---DSDS 707
Db 651 ENQYRKEKEADLLEQORLYADSDS 676

RESULT 3
US-09-722-139-2
; Sequence 2, Application US/09722139
; Patent No. 6355471
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6355471el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1055
; CURRENT APPLICATION NUMBER: US/09/722,139
; CURRENT FILING DATE: 2000-11-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Human
US-09-722-139-2

Query Match 32.5%; Score 1311.5; DB 4; Length 1375;
Best Local Similarity 39.1%; Pred. No. 1.3e-101;
Matches 309; Conservative 117; Mismatches 241; Indels 123; Gaps 13;

QY 6 NIKVVVRPFNFNAREIDRGAKEIVRMENQITLT-----PPPGAEEKARKSGKTMIDGPKA 61
Db 3 SVKVVAVVRPFNFNARETSQDAKCVMSQGNNTSIINP-----KOSKDAKSGFT 51
QY 62 FADRSYWSFDKNA-PNYARQEDLFQDLGVPLLDNAFKYNNCIFAYGTGSGKSYSMGY 121
Db 56 FTYDFSYSADTKSPDVSQEMWFTLGTDDVKSFAEGYNACVAYGTGSGKSYTMGN 115

Query Match		20.2%;	Score 814;	DB 1;	Length 955;
Best Local Similarity		28.1%;	Pred. No. 6.9e-60;		
Matches 253;		Conservative 125;	Mismatches 292;	Indels 230;	Gaps 24;
QY	7	IKVVVVRPFPNARE--IDRGAKCIVRMENQNTILTPP-----PGAEKARKSKTKTMD	57		
DB	13	VKVSVVRPLNERENNAPEGTKVTVAAKQAAVTVVKVLGGSSNNGAAESMGTAARRVAQD	72		
QY	58	GPKAFAPDRSWSF---DKNAPNARQEDLFQDLGVPLLDNAFKYNNCIFAYGQTGSGK	114		
DB	73	----FQDFHVSFVSTPDACGATPATQADVFTIGTGYPLVQHFADGNSCLFAYGQTGSGK	128		
QY	115	SYSMG-----YGEHGVIPRICODMFRINELQDKNLCTCTVEVSYLEIYNERNVRLD	168		
DB	129	TYTMGADVSAISGGGNGVTPRICLEIFARKASVEAQAQHSRWIVELGYVEVYNERNVSDLL	188		
QY	169	NPSTKG-----NLKVRHPSSTGPIVEDLAKLVRSFOELENLMDGCKARTVAATNM	220		
DB	189	GKRKKVGKGGEEVYDVREHPSRGVFLGQRQLVEGSLDDVVRLTEIGNGVRHTASTKM	248		
QY	221	NETSSSHAVFTLTL---TQKHDEETKMDTEKVAKISLVLDLAGSERATSTGATGARKL	276		
DB	249	NKSSSHALIMLLREERTNWKSETIRTAGKSRMNLVDLAGSERVAQAQVQEQQPK	308		
QY	277	EGAEINRSLSLTIGRVTAALADMS--GQKKNQLVPYRDSVLTLKDSLGGSMTAMIAA	335		
DB	309	EATHINLSLTIGRVTDVLADMATKGAQYSVAPFRDSKLTIFILKDSLGGSKTFMIAT	368		
QY	336	ISPADINFEETLSTLYADSAKRIKNHAVVNEDPNARMIRELKEELAAQLRSKQSSGGG	395		
DB	369	VSPALNYEETLSTLYASRARDIVNVAQVNEPRARRIRELEEQMEDMRQM-----	421		
QY	396	GGAGSGGPVEE-----SYPDTPLEK	417		
DB	422	--AGGPAYVSELKKLALLESEAKRAADLQALEREHQNQVQERLLRATEAEKSELES	479		
QY	418	QIVSIQ-----PDATVKMKSAAETVEQL	441		
DB	480	RAAALOEAMTATRRQDKMOALNLRKEOARKERELLKEMAKDALSKVRRKDAEIA	539		
QY	442	NQSEKLYRDLNQTWEEKIAKTEIHKEREAALEELGISIEKGVGPVHSEKEMPHLVNLS	501		
DB	540	SEREKL-----ESTVAQLEREQREVALDAL-----QTHQKLOEALESSE	581		
QY	502	-----DPLLAECVYNIKPGQIFVGNVNODTQAEIRLNGSKILKEHCTFENVNVTVI	555		
DB	582	RTAAERDOLLOQ---LTLEQSERITQSLQSVTVDRERLTR-DLQRTQYEGETELARDVAL	638		
QY	556	PNE-----KAAVMVNGVRIDKPTLRSGYRIILGDFHIFRNFHPEEAPAEQESLLRHS	610		
DB	639	AQEMEARYHAAPVFLQTLLELATEWEDALR-----ERALAEDAAAAE--	682		
QY	611	VTSQSGPAPGRDHRITLSKAGSDADGDSRSDSPLPHERGKDSWFVARRAASAAILGLD	670		
DB	683	-----LDAASASTSQN-----ARESACERLTSLE	705		
QY	671	QKISHLTDELDALFDVQKARAVRRGLVEDNEDSDSSFPVRDKYMSNGTIDNFSLDT	730		
DB	706	QOLRE--SEERAELASOLEATAAKSSAEQDRENTATLEQLQRES-----EARAEL	757		
QY	731	AITWPGTPRSDDDGDALEFFGDKKSKQDASNV-----DVEELRQOQAQMEALKTAK	781		
DB	758	ASQLEATAAA-----KMSAEQDRENTATLEQLQRLDSEERAELASQLESTTAAK	807		
RESULT 7					
US-08-282-845-2					
; Sequence 2, Application US/08282845					
; Patent No. 5719263					
; GENERAL INFORMATION:					
; APPLICANT: Reed, Steven G.					

Query Match	20.2%;	Score 814;	DB 1;	Length 955;
Best Local Similarity	28.1%;	Pred. No. 6.9e-60;		
Matches 253;	Conservative 125;	Mismatches 292;	Indels 230;	Gaps 24;
QY	7	IKVVVRVPFPNARE--IDRGAKCIVRMENQNTILTPP-----PGAEKARKSGKTIIMD	57	
DB	13	VKVSVRPPLNERENNAPEGTKVTVAAKAAAVTVKVLGGSNNSGAAESMGTRARRAQD	72	
QY	58	GKPAFAFDRSTWSF---DKNAPNTARQEDLFQDLGVPLLDNAFKGYNNCIFAYGOTSGGK	114	
DB	73	---FQDFHVSWSVETPDACGATPATQADVFTIGYPLVQFAFDGFCNSCLFAYGOTSGGK	128	
QY	115	SYSMVG-----YKREHGVIPRICODMERRINELQDKNLCTCTVEVSYLEIYNERNVRLD	168	
DB	129	TYTMGADVSAISGSGNGVTPRICLEIFARKASVEAQGHSRWIVELGYVEVYNERSDLL	188	
QY	169	NPSTKG-----NLKVRHPSTGPIVEDLAKLVRSFOELENLMDGCKARTVAATNM	220	
DB	189	GKRKKGKVGSGGEVVDVREHPSRGVFLGQRLVEVGSDDVDVRLIEINGVVRHTASTNM	248	
QY	221	NETSSRSHAVFTLT---TQKWHDEETKMDTEKVAKISLVLDLAGSERATSTGATGARKL	276	
DB	249	NDRSSRSHAIIMLLREERTMTTKSETIRTAGKSSRMNLVDLAGSERVAQSOVEGQOFK	308	
QY	277	EGAEINRSLSLTIGRVTAALADMS--GQKKNQLVPYRDSVLTLKDSLGGSMTAMIAA	335	
DB	309	EATHINLSLTIGRVTDVLDAMTKAKAQYSVAPFRDSKLTFLKDSLGGSMTFMAT	368	
QY	336	ISPADINFEETLSTLYADSAKRIKHAHVVDPNARMIRELKEELAQLRSLQSSGGG	395	
DB	369	VSPSALNYEETLSTLYASRADIVNVAQVNEDPARRIRELEEQMEDMRQAM-----421		
QY	396	GGAGSGGPGVEE-----STPDPDTPLEK	417	
DB	422	--AGDPAYVSELKKLALLEBAQKRAADLOALEREREHNOVQERLLRATEAEKSELES	479	
QY	418	QIVSIQO-----PDAIVKMKSAEIVEQL	441	
DB	480	RAAALQEMTATRRQADKKMOALNRLKEQARKERELLKEMAKKDAALSKVRRRKDAEIA	539	
QY	442	NOSEKLYDLNQTWEKIAKTEIEHKEREAAELGELGSIIEKGFGVGYHSEKEMPHLVNLS	501	
DB	540	SEREKL-----ESTVAQLEREQEREVALDAL-----QTHORKLQEALESSE	581	
QY	502	-----DPLLAELVYNIKPGQTRGVNQNQDTQAEIFLNGSKILKEHCITFENVVNTIV	555	
DB	582	RTAAERDQLIQ--LTLEQSERLTQSOVVDRLRLTR-DLQRIQYEGETELARDVALCA	638	
QY	556	PNE-----KAAVMVNGVRIDKPTLRSGYRIILGDFHIFRHNPEEAPAEARQESLLRHS	610	
DB	639	AOEMEARYHAAPVPHLOTLLLELATWEDALR-----ERALARDEAAAAA--	682	
QY	611	VTNSQLGSPAPGRHRTLSKAGSDADGDSRSDSPLPHFRKGDSDNFYARREASAILGLD	670	
DB	683	-----LDAAAASTSQN-----ARSACERLTSLE	705	
QY	671	QKISHLTDELDALFDVQKARAVRGLVEDNEDSDSQSGFFPVVRKYNSNGTIDNFSLDT	730	
DB	706	QOLRE--SEERAAELASQLEATAAKSSAEQDRENTLRATLEQLRES-----PARAEL	757	
QY	731	AITMPGTPRSDDGDLFFGDKKSKQDASNV-----DVEELROOQQAQMEALAKTAK	781	
DB	758	ASQLEATAAA-----KMSAEQDRENTLRATLEQLRDSERAAELASQLESTTAAK	807	

RESULT 7

us-08-282-845-2

;

Sequence 2, Application US/08282845

;

Patent No. 5719263

;

GENERAL INFORMATION:

;

APPLICANT: Reed, Steven G.

```

Query Match      20.2%; Score 814; DB 1; Length 955;
Best Local Similarity 28.1%; Pred. No. 6.9e-60;
Matches 253; Conservative 125; Mismatches 292; Indels 230; Gaps 24;

QY      7 IKVVVRVPFNARE--IDRGAKCTVRMEGNOTILTPP-----PGAEKARKSGKKTMD 57
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      13 VKYSVRVRPLNERENNAPEGTKVTVAAKQAQAAVVTVKVLGGSNNSGAESMGTARRVAQD 72

QY      58 GKAPAFDRSYWSF---DKNAPYARQEDLFQDLGVPLLDNAFKNQYNNCIFYAGTQSGK 114
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      73 ----FQPDHFVWSVETPDACGATPATQADVFTIGYPLVQHAFDFGNSCLFAYGOTGSGK 128

QY      115 SYSMG-----YCKEHCYVPRICODMFRRLNELQKDKNLCTCTVEVSVLEYINERVRL 168
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      129 TYTMGADVSAISGEGGVTPRICLEIFARKASVEAQGHSRWIVELGVYEVYNERVSDLL 188

QY      169 NPSTKG-----NLKVRHPSTGPVVEDLAKLWRSFQEIENLMDGNKARTVAADNM 220
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      189 GKRRKGVKGGEVYVDVREHPSRGVLEGORLVEVGSDDVRLIEIGNVGRHTASTKM 248

QY      221 NETSSRHAVFTLTL----TQKWHDEETKMDTEKVAKISLYDLAGSERSTATGATGARL 276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      249 NDRSSRSHAIMLLREERTMTTKSETIRTAKSSRMNLVDLAGSERVAQSGVEGQFK 308

QY      277 EGAEINSLSTGLRVIAALADMS--GKQKNQLVPYRDSVLTLLKDSLGNSMTAMIAA 335
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      309 EATHINLSLTGLRVIDVLDMATKGAKAQSVAPFRDSKLTFTLKDSLGNSKTFMTAT 368

QY      336 ISPADINFEETLSRLYADSAKRKNHAVNEDPNARMIRELKELAQLRKLSSGGG 395
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      369 VSPALANYEETLSRLYASRADIVNVAQVNEDPARRIRELEQMEDMROM----- 421

QY      396 GGAGSGGPVEE-----SYPPDTPLEK 417
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Query Match      20.2%; Score 814; DB 5; Length 955;
Best Local Similarity 28.1%; Pred. No. 6.9e-60;
Matches 253; Conservative 125; Mismatches 292; Indels 230; Gaps 24;

QY 7 IKVVRVRFNARE--IDRGAKCIVRMEGNOTILTPP-----PGAEEKARKSKTMD 57
DB 13 VKVSVRVLNERENNAPEGTVTVAQAQAAVTVKVLGGSSNGAAESMGTARRVAQD 72

QY 58 GKAFADFSYNSF---DKNAPNYARQEDLFQDLGVPLLDNAFKGYNCCIFAYGOTGSK 114
DB 73 ----FOFDHVFWSVETPDACGATPATQADVFTIGYPLVQHAFDGFSCLFAYGOTGSK 128

QY 115 SYSMMG-----YGKEHGVIPRICQDMFRINELQDKNLTCTVEYSYLEIYNERVDLL 168
DB 129 TYTMMGADVSALSGEGNGVTPRICLEIFARKASVEAQAQHSRWIVELGYVEVYNERVSDLL 188

QY 169 NPSTKG-----NLKVRHPSTGYPVEDLAKLVVRSFOEINLMDGNKARTVAATNM 220
DB 189 GKRRKGKGGGEEVYVDVREHPSRGVFLQGRVLEQSLDDVVRLEIENGVRHTASTKM 248

QY 221 NETSSRSHAVFTLTL---TQKWHDEETKMDTEKVAKISLVDLAGSERATSTGATGARLK 276
DB 249 NDRSSRSHAIIMLLREERTMTTKSGETIRTAGKSSRMNLVDLAGSERVAQSQVEGQOK 308

QY 277 EGAEINRSLTGRVTAALADMSS-GKQKNQLVPYRDSVLTWLLKDSLGGNSMTAMIAA 335
DB 309 EATHINLSLTGLRVIDLADMATKAKAQYSVAPFRDSKLTFLKDSLGGNSKTFMIAT 368

QY 336 ISPADINFEETLSTRYADSAKIKHNAVNPDPNARMIRELKEELAQLSKLQSSGGGG 395
DB 369 VSPSALNYEETLSTRYASRARDIVNVAQVNEPRARRIRELEEQMEDMRQAM----- 421

QY 396 GGAGSGGPPVEE-----SYPPDTPLEK 417
DB 422 --AGGDPAYVSELKKLALLESEAKRAADLQALERERHNOVQERLLRATEAEKSELES 479

QY 418 QIVSIQO-----PDATVKKMSAEIVEQL 441
DB 480 RAAALQEBTATRQADKMQALNLRKEQARKERELLKEMAKDAALS KVRRRKDAEIA 539

QY 442 NQSEKLYDLNQTWEELKATEETHKREAALELGISIEKFGVPYHSKEMPHLVNLS 501
DB 540 SEREKL-----ESTVAQLERQERREVALDAL-----QTHORKLQEALESSE 581

QY 502 -----DPLLAECILYNIKPGQTVGNVNDQTAIEIRLNGSKILKEHCTENVDNVTVI 555
DB 582 RTAAERDQLQO--LTELQSBRTQLSQVTDRELRTR-DLQRIQYEGYGETELARDVALCA 638

QY 556 PNE-----KAAMVNGVRIDKPTLRSGYRILIGDFHIFRHNHPEAREARQESLLRHS 610
DB 639 AQEMEARYHAAVFFHLQTLLELATWEDALR-----ERALAERDEAAAAE-- 682

QY 611 VTNQLGSPAPGRHRTLSKAGSDGDSRSDSPLPHFRGKSDSWFYARREAAASAILGLD 670
DB 683 -----LDAAASTSQN-----ARESACERLTSLE 705

QY 671 QKISHLTDELDAFPDDYQKARAVRGLVEDNEDSDOSSPPVRDKYMSNGTINDFSLDT 730
DB 706 QQLRE--SEERAAELASQLEATAAKSSAEQDRENTRATLEOOLRES-----BARAAEL 757

QY 731 AITMPTGTPRSSDDGDLFFGDKSKQDASNV-----DVEELRQQAQMEALKTAK 781
DB 758 ASQLEATAAA-----KMSAEQDRENTRATLEOQLRDSERAAELASQLESTTAAK 807

RESULT 9
US-08-428-414A-3
; Sequence 3, Application US/08428414A
; Patent No. 5912166
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: LEISHMANIASIS

```

```

; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/428,414A
; FILING DATE: 21-APR-1995
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Kadlecck, Ann T.
; REGISTRATION NUMBER: 39,244
; REFERENCE/DOCKET NUMBER: 210121.407
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANDBERRY
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 955 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-428-414A-3

```

```

Query Match      20.2%; Score 812.5; DB 2; Length 955;
Best Local Similarity 40.0%; Pred. No. 9.2e-60;
Matches 201; Conservative 75; Mismatches 164; Indels 63; Gaps 13;

```

```

QY 7 IKVVRVRFNARE--IDRGAKCIVRMEGNOTILTPP-----PGAEEKARKSKTMD 57
DB 13 VKVSVRVLNERENNAPEGTVTVAQAQAAVTVKVLGGSSNGAAESMGTARRVAQD 72

QY 58 GKAFADFSYNSF---DKNAPNYARQEDLFQDLGVPLLDNAFKGYNCCIFAYGOTGSK 114
DB 73 ----FOFDHVFWSVETPDACGATPATQADVFTIGYPLVQHAFDGFSCLFAYGOTGSK 128

QY 115 SYSMMG-----YGKEHGVIPRICQDMFRINELQDKNLTCTVEYSYLEIYNERVDLL 168
DB 129 TYTMMGADVSALSGEGNGVTPRICLEIFARKASVEAQAQHSRWIVELGYVEVYNERVSDLL 188

QY 169 NPSTKG-----NLKVRHPSTGYPVEDLAKLVVRSFOEINLMDGNKARTVAATNM 220
DB 189 GKRRKGKGGGEEVYVDVREHPSRGVFLQGRVLEQSLDDVVRLEIENGVRHTASTKM 248

QY 221 NETSSRSHAVFTLTL---TQKWHDEETKMDTEKVAKISLVDLAGSERATSTGATGARLK 276
DB 249 NDRSSRSHAIIMLLREERTMTTKSGETIRTAGKSSRMNLVDLAGSERVAQSQVEGQOK 308

QY 277 EGAEINRSLTGRVTAALADMSS-GKQKNQLVPYRDSVLTWLLKDSLGGNSMTAMIAA 335
DB 309 EATHINLSLTGLRVIDLADMATKAKAQYSVAPFRDSKLTFLKDSLGGNSKTFMIAT 368

QY 336 ISPADINFEETLSTRYADSAKIKHNAVNPDPNARMIRELKEELAQLSKLQSSGGGG 395
DB 369 VSPSALNYEETLSTRYASRARDIVNVAQVNEPRARRIRELEEQMEDMRQAM----- 421

QY 396 GGAGSGGPPVEEPPDTPLEKQIVSIQOQDPAATVKKMSAEIVEQLNQLRDLNQTW 455
DB 422 --AGGDPAYVSE-----LKKLALLES-----EAQKRAADLQALERE-----REHNOVQ 463

QY 456 EEKLAKEETHEHKE---REAALEE 475
DB 464 ERLLRATEAEKSELESRAAALQE 486

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: H.sapiens
US-09-541-782-10

Query Match      16.7%  Score 673;  DB 4;  Length 1057;
Best Local Similarity 27.1%  Pred. No. 6.4e-48;
Matches 245;  Conservative 146;  Mismatches 290;  Indels 224;  Gaps 34;

Qy  4  GGNKVVVRPPFNAREIDRGAKCIVRMENQOTILTPPPGAEERKARKSGKTIIDGPKAFAPD 63
Db  16  GKNIQVVYRCRPFNLAERKASAHISVECD-----EVRKEVSVRTGLADKSSRKTYT 67
Qy  64  FDRSYWSDFNAPYARQEDFODLGVPLLDNAPKGYNCCIFAYGOTGGSGKSYNMW--- 120
Db  68  FDMVEGASTK-----QIDVRSVVCPIDEVINGYCTIFAYGOTGTGKTFTMEGERS 120
Qy  121  ----YGKEH-----GVIPRICQDMFRRINELQKDKNLCTVEYSYLEIYNERVRLDLPNST 172
Db  121  PNEEYTWEEEDPLAGIIPRTLHQIPEKLT---DNGTEFSVKVSLLEYNEELFDLLNPSS 176
Qy  173  --KGNLKVREHP--STGPYVEDLAKLVRSFQBIENLMDGKNKARTVAATNNNETSSRS 228
Db  177  DVSERLQMFDDPRNKRGVIIKGLEEITVHNKDEVYQILEKGAARKTTAATLNNAYSSRS 236
Qy  229  AVFTLTLLQKHDEETKMDTE---KVAKISLVDLAGSERATSTGATGARLKEGAGINRSL 285
Db  237  SVFSVTI---HMKEETTIDGELVKIGKLNLDLAGSNIGRSGAVDKRAREAGINOSL 292
Qy  286  STLGRVIAALADMSSGKQKNQLVPRDSVLTLLKDSLGGNSMTAMIAAISPADINFE 345
Db  293  LTGLRVITAVE-----RTPHPVYRESKLTIRIQLDSLGGTRTSIIATISPASLNLE 345
Qy  346  TLSTLRYADSARKRIKHAHVNEDPNAR--MIRELKEELAEQLAKSKLOSSGGGGGAGSGGP 404
Db  346  TLSTLEYAHRKINILNKPEVNQKLTAKALIKETYEIERLKRDLAAAREKNG----- 397
Qy  405  V---EESY---PPDTPLEKOIVSIQQPDATVKKSKAEIVEQLNQSEKLYRD---LNQ 453
Db  398  VYISENFVMSGKLTVOEQIVEL-----IEKGAVE--EELNRVTELFMDNKNELDQ 449
Qy  454  TWEELAKTEIHKEREAALE--ELGISIEKGFVGPYHSKE-----MPHLVNLSDPLLA 506
Db  450  CKSDLQNKIQLETTQKHLQETKQLQVKEEYTSALESTEELKHLDAASKLLNTVEETKD 509
Qy  507  ECLVYNIKFGQTRVGNVNDTQ-----AELRLNGSK-----ILKEHCT--PEN 547
Db  510  VSLGHSKLCRKKAVDQHNAAEQDIFGKNLNSLFNNMEELIKDGSQKAMLEVHKTLFGN 569
Qy  548  -----VDNVVTIVPNEKAAMVGVNRIDKPTRLRGYRIILGDFHIFREHNHPEARA 599
Db  570  LLSVSSALDITTYVALGSLTSIPEN---VSTHVSQIFNMIL----- 608
Qy  600  ERQEOSLLRHVSVT-----NSQLGSPAGRHDRTL 628
Db  609  --KEQSLAAESKTVLQELINVLKTLSSLEMLSPVTVVSIKINSQL-----KHIFT 660
Qy  629  SKAGSDADGDSRSDSPLPFRGKSDSDFEYA-----RREAASAILGLDQKISHLTDD 679
Db  661  SLTVADKTEDQKK-----RNSDGLFSLICNLNHELQENTICSLVESQKQCNLTED 711
Qy  680  -----ELDALFDVQK--ABAVRRGLVEDNEDSDSQSFPVR 714
Db  712  LKTIKQTHSQELCKLMLNLTWERFCALEKCNIIQKPLSVQENIQOKSDIVNKKTFH--S 770
Qy  715  DKYMSNGTIDNFSLDTATIMPTGTPRSDDGDALFFGDKSKQDASNDVDEELRQQAQME 774
Db  771  QKFCADS--DGFSQEL-----RNFQEGTKLVEESVKHSDKLGKLNLEKISQTEQRC 820
Qy  775  EALKT 779

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: H.sapiens
US-09-541-782-10

Query Match      18.5%  Score 747;  DB 4;  Length 357;
Best Local Similarity 46.6%  Pred. No. 6.4e-55;
Matches 170;  Conservative 54;  Mismatches 119;  Indels 22;  Gaps 8;

Qy  6  NIKVVVRPPFNAREIDRGAKCIVRMENQOTILTPPPGAEERKARKSGKTIIDGPKAFAPD 65
Db  3  NQVAVVRPLSKXXXXXXXXXXXXXXXXXXXXVNDNRDPDGFDSREKVM-----AFGFD 58
Qy  66  RSYWSDFNAPYARQEDFODLGVPLLDNAPKGYNCCIFAYGOTGGSGKSYNMWGYGKEH 125
Db  59  YCYSVNPEDEQYASQDVVFDLGEVLGSAKGNICLFAFGTGGSKTYMLGTPTASV 118
Qy  126  GVIPRICQDMFRRINELQKDKNLCTVEYSYLEIYNERVRLDLPNSTKG---NLKVR 179
Db  119  GLTPRICEGLFVR---EKDCASLPSSCRIRKVSFLFIYNERVRLDKSGQKSKYTLVR 174
Qy  180  EHPSTGPTVEDLAKLVRSFQBIENLMDGKNKARTVAATNNNETSSRSRAVFTLTLLTKW 239
Db  175  EHPGPTVQGLSQHVVTNYKQVLIQLEEGIANRITAATHVHEASSRSHAFTIHYTQA- 233
Qy  240  HDEETKMDTEKAKTSLVDLAGSERATSTGATGARLKEGAEINRSLSTLGRVIAALADMS 299
Db  234  -ILENNLSEMAKINLVDLAGSERADPSYCKD--RIAGANINKSLVTGLVISTLAONS 291
Qy  300  ---SGKQKN--OLVPRDSVLTLLKDSLGGNSMTAMIAAISPADINFEETLSTLYAD 354
Db  292  QVFSSQSLNSSSYIPYRDSVLTLLKDSLGGNSKTIINVATVSPAHTSYSETMTSLRYAS 351
Qy  355  SAKRI 359
Db  352  SAKNI 356

RESULT 11
US-09-541-782-10
; Sequence 10, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541.782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
```

```
Db 821 ESLNT 825
|||
RESULT 12
US-09-572-191-2
; Sequence 2, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572,191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-572-191-2

Query Match 16.5%; Score 664.5; DB 4; Length 1388;
Best Local Similarity 27.4%; Pred. No. 5.2e-47;
Matches 258; Conservative 129; Mismatches 327; Indels 227; Gaps 35;

QY 4 GGNKVVVRVPPNAREIDRGAK---CIVRMEGNQITLTPPPGAEKARKSGKTMDGPK 60
| ||| ||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 24 GDAIKVVRIRPPAERSGSGADGEQNLCLSVLSTSLRLHNSP-----EPK 68
| ||| ||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 AFADRSYNSFDKNAPYARQEDLFQDLGVPLDNLAFKYNNCIFAYGQTGSGKSYMMG 120
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 69 TTFDHFV-----ADVDTTQESVFATVAKSIVESCMSGYNGTIFAYGQTGSKTFTMG 121
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 YGKE-----HGVIPRICODMFRIN-ELQK---DNLTCTVEVSLEYIYNERVRDLN 169
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 PSESDNFSHNLGRVIPRSFEYLSLIDREKEKAGAKSFLC--KCSFIEIYNEQIYDLD 179
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 PSTKGNLKYREHSTGPYVEDLAKLVRSQEIENLMDGKNKARTVAATNMNETSSRSA 229
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 SASAG--LYLREHIKKGKGVVAVQVTSAAEAYQVLSGGWRNRVASTSMNRESSRSA 238
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 230 VFTILTQKHDEETKMDTEKVAKISLVDLAGSERATSTGATGARLKEGAEINRSLTGL 289
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 VFTITTESM--EKSNEIVNIRTSLNLVDLAGSERQKDTHAEGMRLKEAGNINRSLSLG 296
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 290 RVTAALADMSGGKQKNQVLPYRDSVLTLLKDSLGNSTMTAAISPADINFEETLST 349
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 QVITALVDVNGKQRH---VCYRDSKLTLFLRDSLGNKAKTAIIANVHPGSRFCGETLST 353
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 350 LRYADSASAKRIKNAHVVEDPNARM-----IRELKEAQLRSLKSSGGGGGAGSGG 403
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 354 LNFQAARAKLIKNAVVEDTQGNVSQLQAEVKRLKEQLAEIAS-----GQT 399
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 404 PVEESYPPDPLEKQIVYSIQPDATVK--MSKAEIYEQLNQSEKLYRDLNQTWEERLAK 461
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 PPESFLRDKKKNYMEYFQGMFLPKKSEQEKSLIEKVTQLE---DLTLKKEPIQS 455
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 462 TEETHKERE---AALEELGISIEKFGVPYHSEMPHLVN-----LSDDDLAECLV 510
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 456 NKMIVKFREDQIIRLEKLHSEKRGFLPEEQDRLLSELRLNEIOTLRQIEHHPRVAKYAM 515
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 511 YN-----IKGQTRVGNVNDTQAEI-----RLNG-----SKILKE 541
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 516 ENHSLREENRLRLLEPVK-RAQEMDAQTIAKLEKAFSEISGMEKSKDNQOGFSPKQAKE 574
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 542 HCTENYD-----NVVTIVPNKEAAV-----MVNGVRIDKPTRLRSYRIILGDPHIF 589
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 575 PCLFANTEKLAQQLIQOTELNNSKQYEBEKEKELTRKQLELESELQS-----LQKANLN 629
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Query Match 15.9%; Score 640.5; DB 4; Length 409;
Best Local Similarity 39.2%; Pred. No. 7.8e-46;
Matches 167; Conservative 62; Mismatches 134; Indels 63; Gaps 12;

QY 4 GGNKVVVRVPPNAREIDRGAK---CIVRMEGNQITLTPPPGAEKARKSGKTMDGPK 60
| ||| ||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 24 GDAIKVVRIRPPAERSGSGADGEQNLCLSVLSTSLRLHNSP-----EPK 68
| ||| ||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 AFADRSYNSFDKNAPYARQEDLFQDLGVPLDNLAFKYNNCIFAYGQTGSGKSYMMG 120
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 69 TTFDHFV-----ADVDTTQESVFATVAKSIVESCMSGYNGTIFAYGQTGSKTFTMG 121
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 YGKE-----HGVIPRICODMFRIN-ELQK---DNLTCTVEVSLEYIYNERVRDLN 169
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 PSESDNFSHNLGRVIPRSFEYLSLIDREKEKAGAKSFLC--KCSFIEIYNEQIYDLD 179
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 PSTKGNLKYREHSTGPYVEDLAKLVRSQEIENLMDGKNKARTVAATNMNETSSRSA 229
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 SASAG--LYLREHIKKGKGVVAVQVTSAAEAYQVLSGGWRNRVASTSMNRESSRSA 238
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 230 VFTILTQKHDEETKMDTEKVAKISLVDLAGSERATSTGATGARLKEGAEINRSLTGL 289
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 VFTITTESM--EKSNEIVNIRTSLNLVDLAGSERQKDTHAEGMRLKEAGNINRSLSLG 296
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 290 RVTAALADMSGGKQKNQVLPYRDSVLTLLKDSLGNSTMTAAISPADINFEETLST 349
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 QVITALVDVNGKQRH---VCYRDSKLTLFLRDSLGNKAKTAIIANVHPGSRFCGETLST 353
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 350 LRYADSASAKRIKNAHVVEDPNARM-----IRELKEAQLRSLKSSGGGGGAGSGG 403
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 354 LNFQAARAKLIKNAVVEDTQGNVSQLQAEVKRLKEQLAEIAS-----GQT 399
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 404 PVEESYPPDPLEKQIVYSIQPDATVK--MSKAEIYEQLNQSEKLYRDLNQTWEERLAK 461
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 PPESFLRDKKKNYMEYFQGMFLPKKSEQEKSLIEKVTQLE---DLTLKKEPIQS 455
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 462 TEETHKERE---AALEELGISIEKFGVPYHSEMPHLVN-----LSDDDLAECLV 510
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 456 NKMIVKFREDQIIRLEKLHSEKRGFLPEEQDRLLSELRLNEIOTLRQIEHHPRVAKYAM 515
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 511 YN-----IKGQTRVGNVNDTQAEI-----RLNG-----SKILKE 541
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 516 ENHSLREENRLRLLEPVK-RAQEMDAQTIAKLEKAFSEISGMEKSKDNQOGFSPKQAKE 574
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
QY 542 HCTENYD-----NVVTIVPNKEAAV-----MVNGVRIDKPTRLRSYRIILGDPHIF 589
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 575 PCLFANTEKLAQQLIQOTELNNSKQYEBEKEKELTRKQLELESELQS-----LQKANLN 629
| : ||| : : : : : : : : : : : : : : : : : : : : : : : :
Query Match 13
US-09-572-191-6
; Sequence 6, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572,191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Human
US-09-572-191-6
```

```

; : | | | | | | | | | | : : | | | | | | : |
Db 354 LNFQRAKLKKNKAVNEDTQGNVSQLOAEVKRLKEQLAEL-----ASQTP 400
Qy 404 PVEESY 409
; : | | :
Db 401 PLEHHH 406

RESULT 14
US-09-541-782-2
; Sequence 2, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; TITLE OF INVENTION: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: A. nidulans
US-09-541-782-2

Query Match 15.7%; Score 633; DB 4; Length 1184;
Best Local Similarity 26.7%; Pred. No. 1.8e-44;
Matches 234; Conservative 161; Mismatches 315; Indels 168; Gaps 35;

Qy 6 NIKVVVRFPFNAREIDRGAKCIVRMEGNQITLPPPGAEEKARKSKTI--MDGPKAFA 63
; | | | | | | | | | | : : | | | | | | : |
Db 81 SIHVVRRCGRNREKVENSGVVLQTEG-----VKGTVLSMGNPAVS 124

Qy 64 FDRSYWFDKNAPYARQEDLPQDLGVPLLDNAFKGNINCIFAYGQSGSKSYMMG--- 120
; : | | | | | | | | | | : : | | | | | | : |
Db 125 -NKTY-TFKRVFSAADQITVYEDVPLVITVEMLAGYNCTIFAYGQTGTMTSGDMT 182

Qy 121 -----YKHEGVIPRICODMFRRLNEQKDKNLCTVEVSYLEIYNERNVRDLLNPSTGN 175
; | | | | | | | | | | : : | | | | | | : |
Db 183 DTGLILDNAGIIPRVLSFLAKLADTES-----TWKSPFIYNEELRDLLSAENPK 236

Qy 176 LKV-----REHPTGPGYVEDAKLVRSFOEINLMDENKARTVAATNMNETSSRSHA 229
; : | | | | | | | | | | : : | | | | | | : |
Db 237 LKIVDNEQKKGHMST--LVQGMETIYDSATAGIKLLQOGSHKQVAATKCNLDLSRSHT 294

Qy 230 VFTLTLTQKHDEETKMDTEKV--AKISLVDLAGSERATSTGATGARLKEGAELNRSLS 287
; | | | | | | | | | | : : | | | | | | : |
Db 295 VFTITVNIK-----RTTESGEYVCPGKLNLDVLAGSENIGRSGAENKRATEAGLINKSLT 351

Qy 288 LGRVIALADMSSCKQKNOLPVYRDSVLTWLLKDSLGGNSMTAMIAISPADINFETLL 347
; | | | | | | | | | | : : | | | | | | : |
Db 352 LGRVINALVD-----KSQHIPPRESKLRLQDSLGGRTKTCIIATMSPARNLEETI 404

Qy 348 STLRYADSAKRIKHVYVNEED-PNARMIRELKEALQRLSKLOSSGGGGGGGGGGGPGVE 406
; | | | | | | | | | | : : | | | | | | : |
Db 405 STLDYAFRAKNIRKNKPOINSTMPKNTLLREPTAIEKLKAEIATRHNGVYMS-----V 459

Qy 407 ESYPPDTPLE---KQIVSIQPDATVKKMSKA---EIVEQLNQSEKLYRDLNQTWEEKLA 460
; | | | | | | | | | | : : | | | | | | : |
Db 460 ESY-EEMKMENESRIISEEQ-RAKIESMESLRRHKVQELTLTSK-FNDLKKDNDOTLA 516

Qy 461 ---KTEETHKREAAELGISIEKGFGVPYHSEKMEPHLVNLDPLLAELCVYNIRKPG 517
; | | | | | | | | | | : : | | | | | | : |
Db 517 ALCTNDVLOQTDIVLQNTRAQLBEEMLCRAHEETER--QLQD---VGKGLISTL--GQ 569

Qy 518 TRVGNVQ-----DTQAEIRLNGSKILKEHCTCFENVDNVVITVIPNEKAAMVNGVRIDKP 572
; | | | | | | | | | | : : | | | | | | : |
Db 570 T-VEDINSIQSKLRKAELDATNAELWRASST--EVSVDTKRIDQRYEAFQTHRAKLLET 626
```

```

Qy 573 TRLSGYRIILGDFHPIRFNHPPEARERQOSLLRHVSVTNSQLGSPAPGRHD----- 625
; : | | | | | | | | | | : : | | | | | | : |
Db 627 TSVK-----VNEFIATEISNIERTSDLSYENRSLDAACNKAETSSAHEDMNVLEE 680

Qy 626 -----RTLKAGSDADGDSRSDPLPHFRGKSDWDFARREASAILGLDQK-----I 673
; | | | | | | | | | | : : | | | | | | : |
Db 681 IKDLREEVKVKGEGNLGSLASAAARISEEVIGETQLHSQSLHTSFNNLCKDLKSIFETMA 740

Qy 674 SHLTD--DELDALFDDVOK-----ARAVR-----RGL 698
; : | | | | | | | | | | : : | | | | | | : |
Db 741 THLSEQKNEINRLRAELQSSNRQNIETHKASAHLAQAIEEHVAAEAEREILMSQIKAL 800

Qy 699 VEDNEDSDSQSSFPVRKYKNSGTIDNFSLDLTATMPTGTPRSDDDGDALFGDKKSKODA 758
; | | | | | | | | | | : : | | | | | | : |
Db 801 VEESR---QKQF-ARLRAKIDGVRTEISASGDMLEQATTOHROIDEWVF---KSEQFA 852

Qy 759 SNV-----DVEELRQQAQMEELAKTAKOE 783
; : | | | | | | | | | | : : | | | | | | : |
Db 853 KDVNASKDEIRTKLQNDWEAFDQNSIIRKATESVHKE 890

RESULT 15
US-08-713-815A-4
; Sequence 4, Application US/08713815A
; Patent No. 5830659
; GENERAL INFORMATION:
; APPLICANT: Russell J. Stewart
; TITLE OF INVENTION: ACTIVE MICROTUBULE-BASED
; TITLE OF INVENTION: SEPARATIONS BY KINESINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thorpe, No. 5830659th & Western, L.L.P.
; STREET: 9035 South 700 East, Suite 200
; CITY: Sandy
; STATE: Utah
; COUNTRY: USA
; ZIP: 84070
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: AST Ascentia 900N
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: Word Perfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/713,815A
; FILING DATE: 13-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Alan J. Howarth
; REGISTRATION NUMBER: 36,553
; REFERENCE/DOCKET NUMBER: T3214/U-2202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (801)566-6633
; TELEFAX: (801)566-0750
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acid residues
; TYPE: amino acid
; TOPOLOGY: linear
US-08-713-815A-4
```

Query Match 15.5%; Score 625; DB 2; Length 411;
Best Local Similarity 39.4%; Pred. No. 1.6e-44;
Matches 154; Conservative 67; Mismatches 128; Indels 42; Gaps 11;

```

Qy 6 NIKVVVRFPFNAREIDRGAKCIVRMEGNQITLPPPGAEEKARKSKGTINDGPKAFAD 65
; | | | | | | | | | | : : | | | | | | : |
Db 12 SIKVVRFPRLNDEEERAGSRFVVRP-----PNVNEC-----ISIAKRVYLF 56

Qy 66 RSYWFDKNAPYARQEDLPQDLGVPLLDNAFKGNINCIFAYGQTGSGKSYSM---MCGY 122
; : | | | | | | | | | | : : | | | | | | : |
Db 57 KVF-----KFN-ASQEKVYNEAKSIIVTDVLACVNGTIFAYGQTSSGKTHTMGVIGDS 109
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2002, 05:38:41 ; Search time 44.64 seconds
(without alignments)
1687.590 Million cell updates/sec

Title: US-09-235-416-1

Perfect score: 4030
Sequence: 1 MSGGNIKVVRVPPNARE.....ELRQQAQMEALTKAKEP 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1674	41.5	1150	1	A55289 kinesin-like prote
2	1663.5	41.3	1695	2	A56921 kinesin family pr
3	1593.5	39.5	1584	1	JN0114 kinesin-related pr
4	1590.5	39.5	1584	2	T15822 kinesin-like prote
5	1396.5	34.7	1921	2	T13827 kinesin-73 - fruit
6	1284.5	31.9	1576	2	T29237 hypothetical prote
7	1147	28.5	928	2	T16759 kinesin-like prote
8	1035	25.7	1121	2	T13750 kinesin-related pr
9	1034	25.7	1121	2	T13750 kinesin-related pr
10	860	21.3	699	1	A57107 kinesin-related pr
11	844.5	21.0	742	1	A57107 kinesin-related pr
12	828.5	20.6	742	1	A57107 kinesin-related pr
13	825.5	20.5	786	2	A53939 kinesin homolog KH
14	814	20.2	955	2	A47334 Lckin kinesin-rela
15	800	19.9	701	1	B44259 kinesin-related pr
16	782.5	19.4	1231	2	A54803 microtubule-associ
17	775	19.2	1229	2	T48959 kinesin-like prote
18	772	19.0	928	2	T10164 kinesin heavy chai
19	765.5	19.0	793	2	JC5831 kinesin-related pr
20	755.5	18.7	1225	2	A56514 chromokinesin - ch
21	746.5	18.5	784	1	A55236 kinesin-related pr
22	744	18.5	1226	2	T51617 kinesin-like prote
23	743	18.4	935	2	T51930 kinesin [imported]
24	722	17.9	929	2	T51932 kinesin [imported]
25	708	17.6	672	2	S54351 kinesin osm-3 - Ca
26	701	17.4	1130	2	T21134 hypothetical prote
27	691.5	17.2	963	1	A41919 kinesin heavy chai
28	689.5	17.1	294	2	S38983 kinesin-related pr
29	687.5	17.1	1031	1	A38713 kinesin heavy chai

30	687	17.0	881	2	184737 kinesin heavy chai
31	684.5	17.0	1032	2	138510 neuronal kinesin h
32	684	17.0	967	1	A35075 kinesin heavy chai
33	681	16.9	968	2	T51933 kinesin motor prote
34	678	16.8	1075	2	T45570 kinesin-like prote
35	677	16.8	1027	2	S37711 kinesin-like prote
36	671.5	16.7	843	2	S44868 kinesin heavy chai
37	671.5	16.7	1056	1	G02157 kinesin-like spind
38	668	16.6	2663	1	S28261 centromere protein
39	665	16.5	670	2	T29898 kinesin protein OS
40	663	16.5	332	2	C48835 kinesin-like prote
41	662.5	16.4	2954	2	T14156 kinesin-related pr
42	661.5	16.4	1388	2	T30335 KLP2 protein - Afr
43	649.5	16.1	975	1	A31497 kinesin heavy chai
44	636.5	15.8	1076	2	B84687 probable kinesin-1
45	635.5	15.8	1083	2	T26844 hypothetical prote

ALIGNMENTS

RESULT 1
A55289
Kinesin-like protein KIF1B - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C:Accession: A55289
R:Nangaku, M.; Sato-Yoshitake, R.; Okada, Y.; Noda, Y.; Takemura, R.; Yamazaki, H.; H
Cell 79, 1209-1220, 1994
A>Title: KIF1B, a novel microtubule plus end-directed monomeric motor protein for tra
A:Reference number: A55289; MUID:95094296
A:Accession: A55289
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1150 <NAN>
A:Cross-references: GB:ID17577; MID:9407338; PIDN:BA04503.1; PID:9407339
C:Superfamily: kinesin-related protein KIF1B; kinesin motor domain homology
C:Keywords: nucleotide binding; P-loop
F:6-354/Domain: kinesin motor domain homology <KNOT>
F:97-104/Region: nucleotide-binding motif A (P-loop)

Query Match 41.5%; Score 1674; DB 1; Length 1150;
Best Local Similarity 43.4%; Pred. No. 1,2e+82;
Matches 380; Conservative 137; Mismatches 237; Indels 122; Gaps 19;

QY	4	GGIKVVRVPPNAREIDRGAKCIYRMGNQIILPPGAERKAKSKGKTMDGKARA	63
DB	3	GASVKAIVRVRPPNSKRETSKSCITIOGNGSTSIINPKPK-----APKSPS	51
QY	64	FDSTYSWF-DKNAPNVARQEDLFQDLGVPLLDNAFGYNNCIPAYGQTSKSGSMGYG	122
DB	52	FDSTYSWHSHPEDPCFASQNRVYNDIGKEMLHAFGIVVCIFAYQGTAGKSYTMGNG	111
QY	123	KEHG--VPIRCODMFRRIINELOKKNLCTYEVSTLEYNEVRDLNPSTGKILKVR	180
DB	112	ESQAVIIPOLCELEFEKIND-NCNEMSYSEVSMEYICERVRDLNPKNGKGNLRVRE	170
QY	181	HSGTGYVEDLAKLIVRSQETIEMDEGNKARTVAATNMNTESSSHVFTLTLOKH	240
DB	171	HPLEGYVEDLSKLVATSTYDIALMDAGKARTVAATNMNTESSSHVFTLTLOKH	230
QY	241	DEETKMDTEKVAKISLVLDAGSERATSGATGRLEGAENRSLTGLRVIALADMS	300
DB	231	DPEPTNLTSEKTSKISLVLDAGSERADSGAKGRLEKANINNSLTGLKVISALAEVK	290
QY	301	GROKQOLVPIRDSVLTWLKDSGNSMTAMIAISPADINPEETLSTLRADSARRK	360
DB	291	KKKKTDFPIRDSVLTWLRLRENLGNSRTAWVAALSPADINDETSLTLYADRAKQK	349
QY	361	NHVVVEDNARIRLEKEIQAIRSKLOSGGGGGAGAGSGPVESSPPDPLKKO--	418
DB	350	CNAVINEDEPNALVRELKEIVTRLKDLRAOGLDITITSMGSLT--SSPSSCSLNSVG	407

[illegible]

```

RESULT      2
A56921
K:kinasin family protein KIF1a - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Feb-2001
C:Accession: A56921
R:Okada, Y.; Yamazaki, H.; Sekine-Alizawa, Y.; Hirokawa, N.
Cell 81, 769-780, 1995
A:Title: The neuron-specific kinesin superfamily protein KIF1a is a unique monomeric motor
A:Reference number: A56921; MUID: 95292344
A:Accession: A56921
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1695 <RES>
A:Cross-references: GB:D29951; NID:g976234; PIDN:BAAB0221.1; PID:g976235
C:Superfamily: kinesin-related protein unc-104; kinesin motor domain homology; plectstrin
C:Keywords: nucleotide binding; P-loop
E:6-360/Domain: kinesin motor domain homology <MOT>
E:97-104/Region: nucleotide-binding motif A (P-loop)

```

[illegible]

```

Db 169 REHILGGYVEDLSIAVTSYNDIÖDLMDSCNKPRTVATNMNETSSSHAVFNIIÖK 228
QY 239 WHDETKMDTEKVAKISLYVDLAGSERATSTGATGARLKEGAEINRSLSLGRVIALADM 298
Db 229 RHDETNITTEKSVKISLYVDLAGSERADSTAGKTRLEKGANINKSLTTLÖKVISALAE 288
QY 299 SSG-----KÖKKNÖLVPRDSVITWMLKDSIGNSMTMIAAISPAÖINFEETSLTRYA 353
Db 289 DSGPNKKKKKKKTDFIPRDSVITWMLKREINLGNSRTMMVAALSPADINDETSLTRYA 348
QY 354 DSAKRIKKNHAVNEDPNAPMRTRELKEELAOJRSKÖSSGGGGGAGSGSGVEESYPPT 413
Db 349 DRAQIRONALIIINDPNNKILRELKDEYTRRLDILYAGGLG-----DIT 392
QY 414 PLEKOIVSIÖOPDATVKKMS-----KAEIVÖLNÖSEKLYDLNÖTW 455
Db 393 DMTNALVGM -SPSSSLISASRAVSLSHERILFAPGSEFAIERLKETEKIYELNETW 451
QY 456 EEKLAKTPEIKERPEAALEELGISTEK--GVGPRHSEMHVLYNSDLPLAECLVYNI 513
Db 452 EEKIRRETAIMERPALLAEVGMAREDEGTLGVFSPKRTPHLYNLNDPLMSCELLIYI 511
QY 514 KPQGTIVGNVQDPOAEIRLNGSKILKEHTEEN-----YNNVYIYNEKAAMVNVNGVR 568
Db 512 KDGYTRVGREAEARRÖDLYVLSGHFTKEEHCFTFRSOSRGCGAAYVTLEFCBQADYVYNNKK 571
QY 569 IDKPTRLNSGYRIILGDFHIFRFNHPPEARAEARÖEÖSLIRHSVTNSÖLSGSPAPGRHRTL 628
Db 572 VTPESTILRSGRRIILMGKSHVFRFNNPEQARÖER----- 606
QY 629 SKASGDADGDSRSPLPHFGKOSDWFYARREASAILGIDÖKISHLTDELDALFDDV 688
Db 607 -----TPCAETPAEPYDMAFAORELLEK -QGDIMK--QEMÖRLQETLEDOY 649
QY 689 OKARAVRGILEVDENDSDOSSFPYRDKYMN 720
Db 650 RREREATYILLE -QÖRLDYESKLEALÖKÖMÖS 680

```

RESULT 3
JN0114
kinesin-related protein unc-104 - *Caenorhabditis elegans*
C/Species: *Caenorhabditis elegans*
C/Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C/Accession: JN0114
R:Otsuka, A.U.; Jeyarakash, A.; Garcia-Araveros, J.; Tang, L.Z.; Fisk, G.; Hartshorne
Neuron 6, 113-122, 1991
A>Title: The C. elegans unc-104 gene encodes a putative kinesin heavy chain-like protein
A/Reference number: JN0114; MUID:91097805
A/Accession: JN0114
A/Molecule type: mRNA
A/Residues: 1-1584 <OTS>
A/Cross-references: GB:M58582
A/Note: 598-Thr and 930-Met were also found
C/Genetics:
A/Gene: unc-104
C/Superfamily: kinesin-related protein unc-104; kinesin motor domain homology; pleckstrin
C/Keywords: ATP; microtubule binding; nucleotide binding; P-loop
F:4-353/Domain: kinesin motor domain homology <KMOT>
F:93-100/Region: nucleotide-binding motif A (P-loop)
F:1285-1287/Region: cell attachment (R-G-D) motif
F:99/Binding site: ATP (lys) #status predicted

```

Query Match      39.5%; Score 1593.5; DB 1; Length 1884;
Best Local Similarity 44.7%; Pred. No. 43=78;
Matches 356; Conservative 139; Mismatches 213; Indels 89; Gaps 19;

QY 6 NIKYVYRFRPNFARIDGACIVAMEGNOTILRPPGAEKARSSGRTIMDGPAAFD 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   3 SVKAAVRRPFPNORISTSTSKVLDDVNNNTTT-----NGSHINRENSFEND 50
Db

YY 66 RSYWSEFDKNAENYARQEDLPDLGVPLLDNAEFKYGNNCIFAGYGTGSGSKYSMMGYGK-- 123

```


A: Molecule type: mRNA
A: Residues: 1-1921 <LH>
A: Cross-references: EMBL:U01788; NID:q1906595; PID:q1906596; PIDN:AA050404.1
C: Genetics:
A: Cross-references: FlyBase:FBgn0019968

Query Match 34.7%; Score 1396.5; DB 2; Length 1921;
Best Local Similarity 48.8%; Pred. No. 2.5e-67;
Matches 305; Conservative 88; Mismatches 171; Indels 61; Gaps 8;

```
QY 7 IKVAVRPFARFARIDGAKCIYMEGNQTLTPPPGAEKARKSGKTINDGPRAPAFDR 66
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 6 IKAVARPFARFARIDGAKCIYMEGNQTLTPPPGAEKARKSGKTINDGPRAPAFDR 58
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 67 SYMSFDKNAPYARQEDLFODLGYPLLDNAFKYNNCIFAYAGTSGSKYSMMGKGEH 126
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 59 CFYSLNDEDEFASQEVYFPCDVGRIIDNMFQGTNACIFAGTSGSKYSMMGKGEH 118
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 127 VIPRICDMERRIWELOKKNLCTVEVSYLIEYNERVRLDNPSTGN-LKREHPSGTG 185
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 119 IIPRLCQDLSALIN-KSTPELMATKVEVSYWEIYNEKHYHDLDPKPKNSLKVREHVMG 177
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 186 PYVEDLAKIVRSFOELTENMDEGNKARTYATMTNETSSSHAVFTLLTLQKHDEETK 245
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 178 PYVDGLSLAVTSYQDIDNLTMEGKSKRTYATMTNETSSSHAVFTLLTLQKHDEETK 237
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 246 MDTEKAKISLVLDAGSERATSTGATGARLKEGAENRSITLGRVIAALADSSGKOK 305
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 238 VSGEKVSRMSLVLDAGSERAVKTCAGVDRLKEGSKNINSLTTLGLVLSKLADSGKSKG 297
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 306 N-QLVPRDSVLTWMLDLSGNSMTAMIAISPADINFEETSLTKYADSAKRIKHA 364
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 298 NDKVPRDSVLTWMLDLSGNSMTAMIAISPADINFEETSLTKYADSAKRIKHA 357
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 365 VNEDPNARITREKEELAQRLSKLSSGCGGAGSGGPGVEESYPPDTPLEKQIVSIQ 424
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 358 VNEDPNARITREKEELAQRLSKLSSGCGGAGSGGPGVEESYPPDTPLEKQIVSIQ 389
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 425 PDATVKKMSAEIYEQINQSEKLYRDLNQTWEERLAKTEELHKEAREALBELGISIEKGF 484
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 390 -----GDVQDKLAESENLKKQISQIWEKLVKTERIQNRQALKEKMGISVQAS- 438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 485 VGPYHSEKEMHLYNLSDPLLAELGYNINIKPGQTRGVNNOVDQAEIRLNGSKLKEHCT 544
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 439 -GKVEKNKXYLVNLDPSLNLVLYLTKDRLLIGRTISGQPPDLSGLGIQPEHCY 497
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 545 FEWVDNVVTIVPEKAAVWNGVRIIDKPTRLSGYRIILGDFHIFRNHPE----- 595
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 498 ITTEDSGLYMEPVQAGCFVNGSAVAEKTPLQNGDRILKGNHFFRVNSPFSKNTSMCAS 557
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 596 -EAREERQOSLLRHSVTSQSLG 618
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 558 EPQTPAQILIDYNFARDEIMONELSN 582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 6
T29237
Hypothetical protein F56E3.3 - *Caenorhabditis elegans*
C: Species: *Caenorhabditis elegans*
C: Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C: Accession: T29237
R: Gattung: S.; Wu, X.
Submitted to the EMBL Data Library, November 1995
A: Description: The sequence of *C. elegans* cosmid F56E3.
A: Reference number: Z20592
A: Accession: T29237
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-1576 <GAT>
A: Cross-references: EMBL:U41536; PIDN:AA052613.1; GSPDB:GN00028; CESP:F56E3.3
A: Experimental source: strain Bristol N2; clone F56E3
C: Genetics:

A: Gene: CESP:F56E3.3
A: Map position: X
A: Introns: 23/1; 47/2; 125/3; 162/2; 191/3; 274/2; 296/3; 498/3; 532/1; 651/3; 723/2;

Query Match 31.9%; Score 1284.5; DB 2; Length 1576;
Best Local Similarity 38.2%; Pred. No. 2.1e-61;
Matches 323; Conservative 120; Mismatches 217; Indels 185; Gaps 23;

```
QY 7 IKVAVRPFARFARIDGAKCIYMEGNQTLTPPPGAEKARKSGKTINDGPRAPAFDR 66
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10 KVAVRPFARFARIDGAKCIYMEGNQTLTPPPGAEKARKSGKTINDGPRAPAFDR 57
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 67 SYMSFDKNAPYARQEDLFODLGYPLLDNAFKYNNCIFAYAGTSGSKYSMMGKGEH 126
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 58 SFGCTDHSYDFASQEVYFPCDVGRIIDNMFQGTNACIFAGTSGSKYSMMGKGEH 117
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 127 VIPRICDMERRIWELOKKNLCTVEVSYLIEYNERVRLDNPSTGN-LKREHPSGTG 185
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 118 IIPRVCHDIFRIQE-TSNSLSFKVEVSYWEIYNEKHYHDLDPKPKNSLKVREHVMG 176
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 186 PYVEDLAKIVRSFOELTENMDEGNKARTYATMTNETSSSHAVFTLLTLQKHDEETK 245
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 177 PYVDGLSLAVTSYQDIDNLTMEGKSKRTYATMTNETSSSHAVFTLLTLQKHDEETK 236
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 246 MDTEKAKISLVLDAGSERATSTGATGARLKEGAENRSITLGRVIAALADSSGKOK 282
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 237 FSGEKVAKISLVLDAGSERAVKTCAGVDRLKEGSKNINSLTTLGLVLSKLADSGKSKG 296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 283 -----RSITLGRVIAALADSSGKOKNOVLPYRDSVLTWMLDLSGNSMTA 331
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 297 VYVPRDSVLTWMLDLSGNSMTAMIAISPADINFEETSLTKYADSAKRIKHA 353
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 332 MIAISPADINFEETSLTKYADSAKRIKHA VNEPNARITREKEELAQRLSKLSSG 391
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 354 MIAISPADINFEETSLTKYADSAKRIKHA VNEPNARITREKEELAQRLSKLSSG 407
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 392 GCGGGAGSGGPGVEESYPPDTPLEKQIVSIQPDATVKKMSAEIYEQINQSEKLYRDL 451
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 408 -----MOTQTKREHATE-----ELERLAESEKRVLAQM 437
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 452 NOTWEERLAKTEELHKEAREALBELGISIEKGFVGPYHSEKEMHLYNLSDPLLAELCY 511
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 438 NKSWEERLAKTEELHKEAREALBELGISIEKGFVGPYHSEKEMHLYNLSDPLLAELCY 495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 512 NIKPGQTRGVN-----VNOPT-----QAEIRLNGSKLKEHC--TFEN 547
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 496 YIN-GSAITIGNSEELTSRDSGLSMTCSDSSRRDDKERTSYLRGLGIMRRHAKMTVEE 554
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 548 VDNVVT--IVP-NEKAAVWNGVRIIDKPTRLSGYRIILGDFHIFRNHPEAREAROEQ 604
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 555 YGGRRLRFLVAPMSSECRICVNGKQITERLLRNGNRLVGMNHFKNCP---KYMDMEQ 611
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 605 SILRHSVTSQSLGPAVGRHDLTKRAGSDADGDSRDSFLPHFRKDSDFYARREAS 664
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 612 SIMEDSTMDY-----NDAMEHVNDANPI-----SS 637
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 665 AILGLDQKISHLDEDELALFDVQKARAVRGVLVEDNEEDSDS--OSSFPVRDKYMSNGT 722
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 638 AV--DOYMSVY-----LKHQEKKAALRQYFAFEKTIOSLT 673
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 723 IDNFSLDTAITMQ-----TPRSSDDGDALEFGKKSKODASNVDELRQQAQMEBALK 778
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 674 AGGTPSTPMT-PGFCILPTITPTGILPPPPANPKQSVSKFFWQAQKEEMFASLTK 722
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 779 TAKOE 783
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 733 RLKAD 737
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 7
T16759
Hypothetical protein R144.1 - *Caenorhabditis elegans*

C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16759
R:Ravello, T.
Submitted to the EMBL Data Library, March 1995
A:Description: The sequence of C. elegans cosmid R144.
A:Reference number: Z18571
A:Accession: T16759
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-928 <FAV>
A:Cross-references: EMBL:U23515; NID:g746492; PID:g746493; PIDN:AAC46545.1; CESP:R144.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:R144.1
A:Introns: 80/3; 125/1; 261/2; 331/3; 350/2; 516/3; 589/1; 615/2; 638/3; 726/1; 865/3; 9

Query Match 28.5%; Score 1147; DB 2; Length 928;
Best Local Similarity 38.8%; Pred. No.2.7e-54;
Matches 281; Conservative 116; Mismatches 217; Indels 110; Gaps 19;

1 MSGGNIKVVVRPFPNAREIDRGAKCIVRMEGNOTILTPPGAEKARKSGKTIMDGP 60
1 MCKGDSIIIVAVRVPFNDREKTRCKLIVEMPEDETTVIRODKTNDK-----K 48
61 AFAPRSTWSPD-----KNA-----PNTARQEDLFQDLGVPLDIAFKYNNCIPAYGQ 109
49 RTTYHSHSYSHSGFSEKKNGLEPTDPHYADQRFVEEDGRCVLANAMAGYNSCFAYGQ 108
110 TSGSGSYSMGYGKRGHGVPRICODMFRINELQDKNLTCTVEVSYLETYNERVRDILN 169
109 TSGSGSYSTYSGKRNKNGIYPIYCEELFKQIAD-NKKKNQFEVYSAMETICEKRDLS 167
170 --PSTKGNLKVREHDPSTGPEVEDLAKLVRSFOEIENTMDEKNKATVAATNNETSRS 227
168 STPPKGGIKLVREHPKNGEYVNLTPVPNSFKIEAKIEEGTKSRTIAATQMNATSSA 227
228 HAVFTLTLQKHDEETKMDTEKAKISLYDLAGSEPRATSTATGARKEGAEINRSST 287
228 HTIVITFENQK--SSKQAGGTSMKSEINLVDLAGEBROSAACTEGDRKEIVINOSLT 286
288 IGRVIALADMSGKOKNQLVPRDSVLTWLLKDSLGNSMTAMIAISPADINEETL 347
287 IGRVIALHDSQAKSGKKTQIPYDSVLTCLKNALGNSKTIIMAIASPADINEETL 346
348 STLRVYADSAKRIRKHAHVNEDPNAMIRELKEELAQLRSKLOS SGGGGGAGSGGPVEE 407
347 STLRVADRAKSIKTNAVVENQTERALRELRRENLRIQSQIQ-----GGTAG----- 393
408 STPPPTPLEKQIVSTIQDPDAIVKMSKAEIVQLNQSEKLYDLNQTWEETLAKTEEIHK 467
394 -----DASNEEIEK--LRQQLAENQKEMEEMESMOOKTA--EEAK 431
468 EREALLEELGISIEKGFVPHYSKEMPHLVNLSDDPLLAECIVYNNKPKQTVGANNQPT 527
432 HASGSEKVEYEMAKK-----KKMCHLMNLNEDPALNVTVHFLPVEGSEVYVGNKPTSS 483
538 QAEIRLNGSKILKEHCTEEN-VDNVVTIYV-NEKAAYVNGVRIKDPTRLRSRGYRIILGD 585
484 GNFITMGSLIPQHVTLTKNOGNQIHLSPCEDDIDIFINGKRVGHEFQLOQNDRAVFPFG 543
566 FIIFEFNIPPEKRAEROESLLRHVSYNSQLSPAPGRDRRLSKRAGSDADDSRSD--- 642
544 NMLYFNNDPTKGG-----IRTDITYENMAQAEIAQINIAALGNGG--LGGSKRDLTL 593
643 -----SPLPHFRGKSDWFEYARREASATILGDKQIS--HLTRDELDAFLDVOQKARAVR 696
594 EEEELMSTLP-----LYQRANMAATELGRNNAKFEIYVSPDK-----RGLTS 634
OY 697 GLVE 700
DB 635 GLTE 638

RESULT 8
T13750
linesin-like protein 38B - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13750
R:Ohkura, H.; Torok, T.; Tick, G.; Hohnsels, J.; Kiss, I.; Glover, D.M.
J. Cell Sci. 110, 945-954, 1997
A:Title: Mutation of a gene for a Drosophila kinesin like protein, KUP38B, leads to f
A:Reference number: Z17752; MUID:97296456
A:Accession: T13750
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1121 <OKR>
A:Cross-references: EMBL:X10667; NID:g2253151; PIDN:CAA71675.1; PID:g2253152
C:Genetics:
A:Cross-references: FlyBase:FBgn0004374
A:Map position: 2
C:Superfamily: kinesin-related protein Eg5; kinesin motor domain homology

Query Match 25.7%; Score 1095; DB 2; Length 1121;
Best Local Similarity 33.7%; Pred. No.4e-48;
Matches 275; Conservative 131; Mismatches 293; Indels 116; Gaps 21;

1 MSGGNIKVVVRPFPNAREIDRG-AKCIVRMEGNOTILTPPGAEKARKSGKTIMDGP 59
115 VSESNMIVAVRVPFLNALCEFRGQVTNVQVYVGNSENLVQAGSADA-SAGVT----- 168
60 KAPAPRSTWSPDKNAPNTARQEDLFQDLGVPLDIAFKYNNCIPAYGQSGKSYSM 119
169 HFEYSQVYVYSCDPERKNACQAKVEGTARPILDAFEGYNACTLPAAYGQSGKSYSM 228
120 GY-----GKEH---GVIPRICODMFRINELQDKNLTCTVEVSYLETYNERVRD 166
229 GIEALDAAIDGPRPDEAGIIPRCHHELFRLEAVKSSQQLQVEVSYFELYNKIND 288
167 LTN-----PSTKGNLKVREHDPSTGPEVEDLAKLVRSFOEIENTMDE 208
289 LLSVGAATAATGSPFIQOQQQQQRPALKVREHPRGVYVVDLSAHVSYSALRWMLAV 348
209 GKARVVAATNNETSRSRHAYF--TLTLQKHDEETKMDTE-----KAKISL 256
349 GNSORATASTAMDKSSRSISFNIYVNLTLSSDTSDDTSYASSLRQTRRSKISL 408
257 VDLAGEBATSTGATGARLKEGAETIRSLSTIGRVIALADMS-----GKOKNQLV 309
409 VDLAGEBATSTGATGARLKEGAETIRSLSTIGRVIALADMS-----GKOKNQLV 309
310 PYRDSVLTWLLKDSLGNSMTAMIAISPADINEETLSTLRVYADSAKRIRKHAHVNEDP 369
469 PYRESVLTWLLRENLGNSKTYMLATISPAIHADTLATLRACARATVNNKVNESR 528
370 NAMIRELKEELAQLRSKLOS SGGGGGAGSGGPVEESTPPPTPLEKQIVSTIQDPDAIV 429
529 HDKIIRDLAEYDRKLSLNEYERORRLSGNSNPNPRKIIET-----SVDETEVRA 581
430 KMSKAEIVQLNQSEKLYDLNQTWEETLAKTEEIHKREALLEELGISIEKGFVGPVH 489
582 LRQQLAERERELSRACK-----SMKEIKLEEDQKSELVYLRKRGGLALEL-----TA 629
490 SKEMPHLVNLSDDPLLAECIVYNNKPKQTVGANNQPT-----VYNOTQAEIRLNGSKILKEHCT 544
630 EOKQACLVNLTADPILISGLTFYLLPQGLVIRIGRILPGSSSSSQPDIVLDGPLVALQHS 689
545 FENV--DNVVTIYVNNKAAVYNGVRIKDPTRLRSRGYRIILGDHFIFERNHP--EEAA 599
690 IEHERGKLYVLPGSEDFETVYNGELDKDRQLEFHCDRLVIGGSHYFRISNPFCSQGRKA 749
OY 600 EROEGSLRH-----SVTNSQLGSPAPGRDRRLSKRAGSDADDSRSDS-----PLPHFR 649

[illegible][illegible]

0Y 221 NERSRSHAVFLTL-----TQKHDEFTKMDTERAKISTYDVLAGESEATSGATGALK 276
 Db 249 NDRSSRSHAIITMLLRERTMTTKSGETITPAKSSRRNNVLAGESEVAOSQVEQDFK 308
 0Y 277 EGAEINRSLSLTGRVIALADMS-GKQKKNOLVYRDSVLTWLLKDSIGGSMATIA 335
 Db 309 EATHNLSITLTIGRVIDLADMATGAKAQSVAFPRSKLTFLIKDGLGNSKTFMAT 368
 0Y 336 ISPADINFEETLSLRYVDSAKRIKNAHVNEDPNARMIRELKEELAOLRSITLSSGGG 395
 Db 369 VSPSALNTEETLSLTRYASRARDIVNAQVEDPRARRIRELEEQMEDRMQM----- 421
 0Y 396 GGAGSGCGPVE-----SYPPPTPEK 417
 Db 422 --AGDPAIVSELKKALLLESEAKRADLQALEREREHNOVERLLRATPAEKSELES 479
 0Y 418 QIVSIQ-----PDATVKKSKAEIVFOL 441
 Db 480 RAALQDEMTATRRQADKQALNLRLKEQARKERELLKMAKKNAALSYPRRRDAEIA 539
 0Y 442 NOSEKLYNDLQNTWEKLAETEEIHKEREALIEIGISIEKGFVGYHSKENPHLVNSD 501
 Db 540 SEREKL-----ESTVAOLEREQEREVALDAL-----QTHQRKIQEALSESSE 581
 0Y 502 -----DLLAECLVYNNIKPGOTRYGNVNOPTQAIRLNGSKILKEHCTFEVNDVATV 555
 Db 562 RTAERDQILOO--LTEIQSEKTSQVYTDREIRLTR-DLQIQEYETELARVALCA 638
 0Y 556 PNE-----KAAVMVGVIRDKPTRLRSGYRIITLGFHIFRFNHPPEARAEKROEOLRHS 610
 Db 639 AQEMBARVHAANVFHLQTLLELATTEMEDALR-----ERALAEHDEMAAE-- 682
 0Y 611 VTNSQLGSPAREHRDRTLSKAGSDADGSDSDPLPHFRGKDSDFYARREPAASAILGLD 670
 Db 683 -----LDAASTSQN-----ARSCERLTSTLE 705
 0Y 671 QKISHLTDELDALFDYDQKAAVRGLVEDNEDSDSSPFRVRYKYSNGTINDFSIDT 730
 Db 706 QQLRE-SEERAELASOLEATRAAKSSAEOQRENTRAILEOQLRS-----ERAAEL 757
 0Y 731 ATTMGTPRSSDDGDALEFGDKKSRQDASNV-----DVELRQQAQMEELAKTAK 781
 Db 758 ASOLEATFAA-----KMSAQOQRENTRAILEOQLRSEERAELASQLESTYAAK 807

 RESULT 15
 B44259
 kinesin-related protein KIF3A - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_Change 19-Jan-2001
 C:Accession: B44259; S27872
 R:Alta, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.
 J:Cell Biol. 119, 1287-1296, 1992
 A:Title: kinesin family in murine central nervous system.
 A:Reference number: A44259; MUID:93077686
 A:Accession: B44259
 A:Molecule type: mRNA
 A:Residues: 1-701 <Alt>
 A:Cross-references: EMBL.D13645; NID:9220469; PIDN:BAA02166.1; PID:9220470
 A:Experimental source: brain
 A:Note: sequence extracted from NCBI backbone (NCBI:118911)
 C:Complex: heterodimer with KIF3B (PIR:A57107); the KIF3A/3B heterodimer associates with
 C:Function:
 A:Description: KIF3 complex is a motor protein that provides anterograde fast axonal transport
 A:Description: kinesin-related protein KIF3; kinesin motor domain homology
 C:Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotide
 F:1-368/Domain: head globular #status predicted <HGI>
 F:1-351/Domain: kinesin motor domain homology <KMO>
 F:100-107/Region: nucleotide-binding motif A (P-loop)
 F:369-559/Domain: helical rod #status predicted <ROD>
 F:600-701/Domain: tail globular #status predicted <TGL>
 F:106/Binding site: ATP (lys) #status predicted

Query Match	19.9%	Score 800	DB 1	Length 701
Best Local Similarity	31.8%	Pred. No. 1e-35		
Matches 231	Conservative 110	Max. Matches 231	Indels 154	Gaps 25
QY	6	NKVVVVRVPPFARAREID---RGAKCIVRREGNOTILTPPGAEAKARKGKGTIMDOPKAF	62	
DB	14	NKVVVVRCPPLNEREKSMKYROAVSYDEKRGITV-----HKTDSN-----EPPKTF	61	
QY	63	AFDRSYWSEFDKNAIPYARQEDLFODLGVPLLDNAEKGYNCCIFAYGQGTSGKSYSMGYG	122	
DB	62	FEDTJVF-----GPE-SKOLDVYNTLARIIDISVLEGYNGTIFAYGQGTGKTFMEGVR	114	
QY	123	KE---HGVIPIRCODMFRRINELQDKNLCTVEVSYLEYIERVRADLLNPSTKGLKLR	179	
DB	115	AVPGLRGVPIPSFAHIFGHIAAEGDTR--LVRSYLEYIEVEVDLLGKDOTORLEVK	172	
QY	180	EHPSTGPVYEDLAKLVRSFOEIEIMDGNKARFYAATNMNSETSSRSYHAFVLTU--TQ	237	
DB	173	KRPDGVYIKDLASAYVYVNNADMDMRIMTLGHKNRSYGATNMNHSRSYHAFITITCECE	232	
QY	238	KKHDEETKMDTERVAKISLVLDLAGSERATSTGATGARLKEGAEINRSISTLGRVIALAD	297	
DB	233	KQVDNRMHY---RMGLIHLVLDLAGSERQAKTGATGCRLEAKYKINISLTLGNVISALVD	289	
QY	298	MSSGKKKKQLVTPYRDSVLTWLLKSLGNSMTAMIAALSPADINFEETLSTIRYVDSK	357	
DB	290	-----GKSTHVPYRNSKTLRLDSDLGNSKTMKCANIGPADYINDETISTLRVIANRAK	343	
QY	358	RKKNHAVNEDENARNIRELKEELQALRSKLOSSGG-GGGAGGSGGCPPEE-----	407	
DB	344	NIKKNARINEDPKDALLRQFOKEIEELKKKLEGESEVSSDISSEDDDEGELGIDGKK	403	
QY	408	-----SYPPDTPLEKQ-----IYS	421	
DB	404	KKRRDQAKKKVSPKKNVEMQAKIDEERKALETKIDMEEBERKNKRAELERRERKDLKA	463	
QY	422	IQDPATVATKMSKAAE-----IYEQUNOSEKLYRDINQOTWEETLAKTEIIRKREKA	472	
DB	464	QOEHQSLEKLTALAEKKVYVGGVDLLAKAEQEKLEEBENMELEERARRAEOURLKELEK	523	
QY	473	-----LEELGISIEKGFGVYHKSKEPMLVNLSDPDLAECVLVNIKPGQTRVGNVND	526	
DB	524	EOERLDIEEKYYSLOEAAQ--KTKLKKRVWTY-----LMAKSEMAIDLQOB	568	
QY	527	TOAEIR--LNGSKILKEHETFEF--VDNVY-----TIYVN-----EKANVY	564	
DB	569	HORELEGLEENITRQLSRELRLQMLIIDNTIPDOYOMIENYVHWNEDIGEMOLKCYAAYG	628	
QY	565	NGCVRIADKPT---RLNSGRIILGDFHIFPNNPEEARAEQEOSILRHSTVNSLOGLSP	621	
DB	629	NNMRKQOTPYDQKKERDPEFVDS--HYV-LATTEEL--RQSLMKLERRTISGKARPAT	683	
QY	622	GRHDP	627	
DB	684	GRKRS	689	

Search completed: July 25, 2002, 05:46:13
Job time: 452 sec

0Y 221 NTSRSAAVFLTL-----TQKHDEFTKMDTERAKISTYDLAGESEATSGATGALK 276
 Db 249 NDRSSSHAIITMLLRERTMTTKSGETITPAKSSRNVLVLAGESEVAOSQVEQDFK 308
 0Y 277 EGAEINRSISTGRIYALADMS-GKQKKNOLVYRDSVLTWLLKDSIGGSMATIA 335
 Db 309 EATHTNLSTITGRIVDLADMATGAKAQSVAFPRSKLTFLIKDGLGNSKTFMAT 368
 0Y 336 ISPADINFEETSTLYVDSAKRIKNAHVNEDPNARMIRELKEELAOLRSITLSSGGG 395
 Db 369 VSPSALNTEETSTLYRSARDIYNAQVEDPRARRIRELEEQMEDRMQAM----- 421
 0Y 396 GGAGSGCGPVE-----SYPPPTPEK 417
 Db 422 --AGSDPAVSELKKALLLESEAKRADLQALEREREHNOVERLLRATPAEKSELES 479
 0Y 418 QIVSIQ-----PDATVKKSKAEIVFOL 441
 Db 480 RAALQDEMTATRRQADKQALNLRLKEQARKERELLKMAKKNAALSYPRRRDAEIA 539
 0Y 442 NOSEKLYNDLQWTEEKLAKTEETIHKEREALIEIGISIEKGFVGYHSKENPHLVNSD 501
 Db 540 SEREKL-----ESTVAOLEREQEREVALDAL-----QTHQRKIQEALSESSE 581
 0Y 502 -----DLLAECLVYNNIKPGOTRYGNVNOPTQAIRLNGSKILKEHCTFEVNDVATV 555
 Db 562 RTAERDQILOO--LTEIQSEKTSQVYTDREIRLTR-DLQIQEYETELARVALCA 638
 0Y 556 PNE-----KAAVMVGVIRDKPTRLRSGYRIITLGFHIFRFNHPEAARAERQOSILRHS 610
 Db 639 AQEMBARVHAANVHLQTLLELATTEMEDALR-----ERALAEHDEMAAE-- 682
 0Y 611 VTNSQLGSPAREHRDRTLSKAGSDADGDSRSDPLPHFRGKDSDFYARREPAASAILGLD 670
 Db 663 -----LDAASTSQN-----ARSCERLTSTLE 705
 0Y 671 QKISHLTDELDALEFDYDQKAAVVRGLVEDNEDSDSSFPVPRDKYMSNGTINDFSIDT 730
 Db 706 QQLRE-SEERAELASOLEATRAAKSSAEOQRENTRAILEOQLRS-----ERAAEL 757
 0Y 731 ATTMGTPRSDDDGALFEFGKRSKODASNV-----DVEELRQQAQMEELAKTAK 781
 Db 758 ASOLEATFAA-----KMSAQOQRENTRAILEOQLRSEERAELASQLESTYAAK 807

 RESULT 15
 B44259
 kinesin-related protein KIF3A - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_Change 19-Jan-2001
 C:Accession: B44259; S27872
 R:Alta, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.
 J:Cell Biol. 119, 1287-1296, 1992
 A:Title: kinesin family in murine central nervous system.
 A:Reference number: A44259; MUID:93077686
 A:Accession: B44259
 A:Molecule type: mRNA
 A:Residues: 1-701 <Alt>
 A:Cross-references: EMBL.D13645; NID:9220469; PIDN:BAA02166.1; PID:9220470
 A:Experimental source: brain
 A:Note: sequence extracted from NCBI backbone (NCBI:P:118911)
 C:Complex: heterodimer with KIF3B (PIR:A57107); the KIF3A/3B heterodimer associates with
 C:Function:
 A:Description: KIF3 complex is a motor protein that provides anterograde fast axonal transport.
 A:Description: kinesin-related protein KIF3; kinesin motor domain homology
 C:Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotide
 F:1-368/Domain: head globular #status predicted <HGI>
 F:1-351/Domain: kinesin motor domain homology <KMO>
 F:100-107/Region: nucleotide-binding motif A (P-loop)
 F:369-559/Domain: helical rod #status predicted <ROD>
 F:600-701/Domain: tail globular #status predicted <TGL>
 F:106/Binding site: ATP (lys) #status predicted

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2002, 05:39:31 ; Search time 25.37 Seconds

(without alignments)
1196.536 Million cell updates/sec

Title: us-09-235-416-1

Perfect score: 4030
Sequence: 1 MSGGNIKVVVRRPFPNARE.....ELRQQAQMEALTKAKQEP 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1667	41.4	1103	1 KFLC_HUMAN	043896 homo sapien
2	1663.5	41.3	1690	1 KFLA_HUMAN	012756 homo sapien
3	1663.5	41.3	1695	1 KFLA_MOUSE	P33173 mus musculu
4	1660.5	41.2	1816	1 KFLB_MOUSE	Q60575 mus musculu
5	1623.5	40.3	1816	1 KFLB_HUMAN	060333 homo sapien
6	1617	40.1	1097	1 KFLD_RAT	Q37877 rattus norv
7	1593.5	39.5	1584	1 U104_CAEEL	P28678 caenorhabdi
8	1554.5	38.6	689	1 KFLB_RAT	088658 rattus norv
9	860	21.3	699	1 K122_STRPU	P48872 strongyloe
10	848.5	21.1	747	1 KFLB_HUMAN	015066 homo sapien
11	844.5	21.0	747	1 KFLB_MOUSE	Q61771 mus musculu
12	827.5	20.5	742	1 K121_STRPU	P46871 strongyloe
13	825.5	20.5	786	1 FL10_CHURE	P46869 chlamydomon
14	814	20.2	955	1 KFLN_LEICH	P46865 leishmania
15	800	19.9	701	1 KFLA_MOUSE	P26741 mus musculu
16	789.5	19.6	702	1 KFLA_HUMAN	Q09496 homo sapien
17	782.5	19.4	1231	1 KFLA_MOUSE	P33174 mus musculu
18	774	19.2	1232	1 KFLA_HUMAN	095239 homo sapien
19	772	19.2	928	1 KFLN_NEUCR	P44667 neurospora
20	770.5	19.1	796	1 KFLC_MOUSE	Q30666 mus musculu
21	769.5	19.1	796	1 KFLC_RAT	051165 rattus norv
22	765.5	19.0	793	1 KFLC_HUMAN	014782 homo sapien
23	755.5	18.7	1225	1 KFLA_CHICK	Q09640 gallus gall
24	746.5	18.5	784	1 KFLB_MOUSE	P46867 drosophila
25	744	18.5	1226	1 KFLA_XENLA	Q91784 xenopus lae
26	743	18.4	935	1 KFLN_SYNRA	Q40933 syncerphala
27	708	17.6	672	1 KFLN_CAEEL	P46873 caenorhabdi
28	691.5	17.2	963	1 KFLN_HUMAN	P33176 homo sapien
29	691	17.1	963	1 KFLN_MOUSE	Q61768 mus musculu
30	687.5	17.1	1031	1 KFLN_STRPU	P35978 strongyloe
31	684.5	17.0	1032	1 KFLN_HUMAN	Q12840 homo sapien
32	684	17.0	967	1 KFLN_LOLPE	P21613 loligo peal
33	676.5	16.8	1027	1 KFLN_MOUSE	P33175 mus musculu

34	673	16.7	1057	1 EGS_HUMAN	P52732 homo sapien
35	671.5	16.7	815	1 KFLN_CAEEL	P34540 caenorhabdi
36	670.5	16.6	957	1 KFLC_HUMAN	Q60282 homo sapien
37	669	16.6	956	1 KFLC_MOUSE	P28738 mus musculu
38	668	16.6	2663	1 KFLC_HUMAN	Q02224 homo sapien
39	651.5	16.2	975	1 KFLN_DROME	P17210 drosophila
40	634	15.7	1056	1 K125_ARATH	P82266 arabidopsis
41	633	15.7	1184	1 B1MC_EMENT	P17120 emericeella
42	631.5	15.7	883	1 KLP5_SCHPO	O14343 schizosacch
43	630	15.6	1067	1 EGS2_XENLA	Q91783 xenopus lae
44	626.5	15.5	1111	1 KFLP_YEAST	P28742 saccharomyc
45	623	15.5	1060	1 EGS1_XENLA	P28025 xenopus lae

ALIGNMENTS

```

RESULT 1
ID KFLC_HUMAN STANDARD; PRT; 1103 AA.
AC 043896; 075186;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kinesin-like protein KIFLC.
GN KIFLC OR KIAA0706.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hippocampus;
RX MEDLINE=98352063; PubMed=9685376;
RA Dorrer C., Closek T., Mueller S., Moeller N.P.H., Ullrich A.,
RA Lammer R.;
RT "Characterization of KIFLC, a new kinesin-like protein involved in
RT vesicle transport from the Golgi apparatus to the endoplasmic
RT reticulum."
RT J. Biol. Chem. 273:20267-20275(1998).
RL [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RT DNA Res. 5:169-176(1998).
RL -I- FUNCTION: MOTOR REQUIRED FOR THE RETROGRADE TRANSPORT OF GOLGI
RL VESICLES TO THE ENDOPLASMIC RETICULUM. HAS A MICROTUBULE PLUS END-
RL DIRECTED MOTILITY.
CC -I- SUBUNIT: MONOMER (POTENTIAL).
CC -I- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED, WITH MOST
CC ABUNDANT EXPRESSION IN HEART AND SKELETAL MUSCLE.
CC -I- PTM: PHOSPHORYLATED ON TYROSINE.
CC -I- SIMILARITY: CONTAINS 1 FHA DOMAIN.
CC -I- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U91329; AAC52117.1; -.
CC DR EMBL: AB014606; BAA31681.1; -.
CC DR HSSP: P17119; 3KAR.
CC DR MIM: 603060; -.

```


Query Match 41.3%; Score 1663.5; DB 1; Length 1690;
 Best Local Similarity 46.7%; Pred. No. 4.7e-81;
 Matches 351; Conservative 125; Mismatches 167; Indels 109; Gaps 15;

```

OY 4 GGNIRVVVRPPNAREIDRGAKCIYRMGNGTILTPPGAEKARKSGKTIIMDGPKARA 63
DB 3 GASVAVAVRPPNAREIDRGAKCIYRMGNGTILTPPGAEKARKSGKTIIMDGPKARA 63
OY 64 PDRSYWSPDKNAP---NARQEDLFDQDGLVPLLDNAFKGYNNCIRAYGOTSGSKSYSMAG 120
DB 52 FQSYWSPDKNAP---NARQEDLFDQDGLVPLLDNAFKGYNNCIRAYGOTSGSKSYSMAG 109
OY 121 YGR--EHGVIPRICODMERINELQKDKNLCTVESEYLEYINERYRDLNLTSTGKNGLV 178
DB 110 KOEKQGGIIPQIGEDLSRINDTND--NMSYSVESEYLEYINERYRDLNLTSTGKNGLV 168
OY 179 REHPTGPEVEDLAKLVVRSFOEINLMDGKAKARTVAATNNETSSRSNAVFTLTQK 238
DB 169 REHPTGPEVEDLAKLVVRSFOEINLMDGKAKARTVAATNNETSSRSNAVFTLTQK 228
OY 229 WHDEETKMDTEVKATISLVLDLAGSERASTGATGARKLGKAGINLSLTGLGVIALAM 298
DB 229 RHDAETNTTEKVSISLVLDLAGSERASTGATGARKLGKAGINLSLTGLGVIALAM 288
OY 299 SSG-----KQKNQOLVPRDSVLTWLLKDSLGNSMTAMIAISPADINEFTLTGRLA 353
DB 289 DSGPKNNKKKKTDFIPYRDSVLTWLLKDSLGNSMTAMIAISPADINEFTLTGRLA 348
OY 334 DSARKIKHNAVVEDPNAMIRELKEELAQLRSKLGSGGGGAGGSGGPGVEESTPYDT 413
DB 349 DRAKQIRCNNAVINEPNKMLIRELKEDEVTLRDLLYAQIG-----DIT 392
OY 414 PLEKQIVSIQODATVKKMS-----KAELVEQUNOSEKLYRDLNQW 455
DB 393 DMTNALVGR-SSSSLSLSSRAASVSSILHERILFAPGSESEIERKEKELIAELNEMW 451
OY 456 EEKLAETIEIKEREALAEELGISIEK--GEVGPYSHKEMPHLVNDDPLAECLVNI 513
DB 452 EEKLAETIEIKEREALAEELGISIEK--GEVGPYSHKEMPHLVNDDPLAECLVNI 511
OY 514 KPGQTRGVNODTOAEIRLNGSKLKEHCTEN-----VDNVVTIVPEKAAVAVNGVR 568
DB 512 KDGIRVGVNODTOAEIRLNGSKLKEHCTEN-----VDNVVTIVPEKAAVAVNGVR 571
OY 569 IDKPTRLSGYRIITGDPIIFRPNPEEARAROEGLSLRHVTSNQLGSPAGRHDRL 628
DB 572 VTEPILNSGNRIIMKSHVFTPEQAKQERER----- 606
OY 629 SKAGSDADODSDSPPLPFRGKDSDFEYARREASALIGLQKISHLTDDDELAFDDV 688
DB 607 -----TPCAETPAEPYDMVAFAQRELEK--QGIDMK--QEMEGRLQGELEDY 649
OY 689 OKARAVRGLVEDNEDSDSGSPVPRDKYMSN 720
DB 650 RREAREATYLE-QORLDYESKLEALQKOMS 680

```

RESULT 3
 KFLA_MOUSE STANDARD: PRT, 1695 AA.
 AC P3173; 061770;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Kinesin-like protein KIF1.
 GN KIF1A OR KIF1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95292344; PubMed=7539720;
 RA Okada Y., Yamazaki H., Sekine-Alizawa Y., Hirokawa N.;
 RT "The neuron-specific kinesin superfamily protein KIF1A is a unique
 RT monomeric motor for anterograde axonal transport of synaptic vesicle
 RT precursors.";
 RL Cell 81:769-780(1995).
 RN [2]
 RP PRELIMINARY SEQUENCE OF 100-247 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93077686; PubMed=1447303;
 RA Alizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M.,
 RA Hirokawa N.;
 RT "Kinesin family in murine central nervous system";
 RL J. Cell Biol. 119:1287-1296(1992).
 CC -1- FUNCTION: MOTOR FOR ANTEROGRADE AXONAL TRANSPORT OF SYNAPTIC
 CC VESICLE PRECURSORS.
 CC -1- SUBUNIT: MONOMER.
 CC -1- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN ADULT BRAIN
 CC TISSUE (MAINLY IN THE CEREBELLUM AND CEREBRUM) WITHIN A SINGLE
 CC TYPE OF NEURONAL CELL. WITHIN THE NEURONAL CELL LEVELS ARE
 CC CONCENTRATED AROUND THE AXON, WITH SMALLER AMOUNTS IN THE
 CC PERINUCLEAR AND SYNAPTIC REGIONS.
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104
 CC SUBFAMILY.
 CC -1- SIMILARITY: CONAINS 1 FHA DOMAIN.
 CC -1- SIMILARITY: CONAINS 1 PH DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: D29951; BAA06221.1; -.
 DR PIR: E44259; E44259.
 DR HSSP: P17119; 3KAR.
 DR MGD: MGI:108391; Kif1a.
 DR InterPro: IPR000253; FHA_domain.
 DR InterPro: IPR001849; PH.
 DR InterPro: IPR001752; kinesin.
 DR Pfam: PF00498; FHA; 1.
 DR Pfam: PF00225; kinesin; 1.
 DR Pfam: PF00169; PH; 1.
 DR PRINTS: PR00380; KINESINHEAVY.
 DR SMART: SM00240; FHA; 1.
 DR SMART: SM00239; KISC; 1.
 DR SMART: SM00233; PH; 1.
 DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE: PS50067; KINESIN_MOTOR_DOMAIN2; 1.
 DR PROSITE: PS50006; FHA_DOMAIN; 1.
 DR PROSITE: PS50003; PH_DOMAIN; 1.
 KW Motor protein; Microtubules; ATP-binding; coiled coil.
 FT DOMAIN 1 361
 FT DOMAIN 366 383
 FT DOMAIN 429 462
 FT DOMAIN 516 572
 FT DOMAIN 622 661
 FT DOMAIN 801 822
 FT DOMAIN 1580 1678
 FT NP_BIND 97 104
 FT SEQUENCE 1695 AA; 191724 MW; D6ECB88CB8C6C6 CRC64;

Query Match 41.3%; Score 1663.5; DB 1; Length 1695;
 Best Local Similarity 46.5%; Pred. No. 4.8e-81;
 Matches 350; Conservative 126; Mismatches 167; Indels 109; Gaps 15;

```

OY 4 GGNIRVVVRPPNAREIDRGAKCIYRMGNGTILTPPGAEKARKSGKTIIMDGPKARA 63

```



```

FT DOMAIN 668 737 COILED COIL (POTENTIAL).
FT DOMAIN 841 869 COILED COIL (POTENTIAL).
FT DOMAIN 1702 1799 PH.
FT NP_BIND 97 104 ATP (POTENTIAL).
FT VASPLIC 289 294 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT VASPLIC 394 434 IDPLIDYSGGKYLDFDNKRRYLASRNORGNSTA
FT VASPLIC 707 1196 IDPLIDYSGGKYLDFDNKRRYLASRNORGNSTA
-> T (IN ISOFORM 2 AND ISOFORM 3).
YESKQALQROVERSLAAETEEEEEEVEEVPWQHEBELA
OMAEKRNKSHQFTSLRDLGMNAVYLAIRKANAISVELKKVQ
FOEVLDTLTSVPPELLEPEMEKTEHEDREPTLVAVEV
ODLKNQATHWASLDLKLORLDMREMDRAGEVASNODDS
ETTWTGSDPTDFRHWKRLVSSPTFHCCVVERLADRTSP
TESTADSDITELADEQDAMEDDEADEADVDGSDAGEEG
SELFSDGHDPEYDRSPFVLIVGRAFYVLSNLYPPLIHRV
AIVSEKGEVRGFLVAVOALADEAPVGGIGRSGTAKI
SEFNEVFNQSDFSANATRSGLSLEELIVGGQOSSSVTI
PPEEKNMNDLDLSCGTLDDCKMVEGSEFIGNIKRIGSA
FTFRVYVLOASGILPEVADICQFNELEHRDEASTEBLKN
NGRSPLEFYHVQNIHAVEVSEFVDTIKTRDIVEVEGH
-> ADSGSDSDSKRSCSESWKLITSLREKLPSRLQTVK
KCGLPSSGKKREPIKMOIIPORRLSKSKSWTTISDLKIQ
VKEICYEVALNDFRHSROEIALAIVKKEICAMGKKDPN
ERDSRAVARDVMDTVGDEKIDMTGTCGGTDVDDLYC
HIDKLEDTLQEVYKONNKDEITLYRKAKMLKMEVYDLIC
SOEOKSOGSHKTEPLVAGANSVDNGSKESGELKEER
VSQLMNGDPARFROGLRMROEQIRFNKLIQOELITQLRQ
NVPHRTIPENRKPRFPKSNPKHRNSPGTHIITDEY
IELRIPKDEESKKEESOKVGAARSDVOSAGTRSD
HIOVSKHISNOOPPOLRMRSNSLNGCOPITTRGOATASS
ESLNSHSHPTADLOTROAKRHIOHRRPYCNYNTGCVBG
STASCCQKQIDPESHQNOFVTPPRMRROFSAPNLKAGRETT
V (IN ISOFORM 3).
MISSING (IN ISOFORM 3).
G -> V (IN REF. 1 AND 2).
GGTL -> RGDI (IN REF. 1).
P -> S (IN REF. 4).
KL -> TW (IN REF. 3).
I -> V (IN REF. 3).
D -> G (IN REF. 4).
SEQUENCE 1816 AA; 204080 MW; E316EC295138E5DE CRC64;

Query Match 41.2%; Score 1660.5; DB 1; Length 1816;
Best Local Similarity 45.1%; Pred. No. 7.6e-81;
Matches 357; Conservative 127; Mismatches 170; Indels 137; Gaps 17;

OY 4 GGNIVVYVRPRENAREIDRGAKCIYRMGNOTILTPPGAEKARKSKTINDGPKARA 63
| : : : | | | | | | | | : : : : : | : : : | | : : : |
DB 3 GASVAVAVRPRENAREIDRGAKCIYRMGNOTILTPPGAEKARKSKTINDGPKARA 51
| : : : | | | | | | | | : : : : : | : : : | | : : : |

OY 64 FDRSYWSE-DKNAPYARQEDLFODLVPLDNARFGYNNCFAYGQTSGRKSYMMGYG 122
| | | | | | | | : : : : : | : : : | | : : : |
DB 52 FDSYWSHSTSPEDPCFASGNRYNYNDIGKEMLHAPEGVVVCIFPAIGQTAGSYTMMQ 111
| : : : | | | | | | | | : : : : : | : : : | | : : : |

OY 123 KEH--GVYPRICQDMFRINELQDKKNTCTVEVSYLEYNERVDLMPSTKGNLKARE 180
| : : : | | | | | | | | : : : : : | : : : | | : : : |
DB 112 BESQAGIIPOLCEELFEKIND--NCNEEMSYSEVSEYMEIYCRVYDILLPKPKGNLRYAE 170
| : : : | | | | | | | | : : : : : | : : : | | : : : |

OY 181 HPSPTPYEDLAKLVYRSFOEIJENLDEGNCARVAAATNMNTESSRSRAVFTLLTLOKH 240
| | | | | | | | : : : : : | : : : | | : : : |
DB 171 HPLCLPYVEDLSKLVAVTSTTDIADLMDAGNKAITYAATNMNTESSRSRAVFTLVFTOKO 230
| : : : | | | | | | | | : : : : : | : : : | | : : : |

OY 241 DEETKMDTEKVAKISLVDLASGERATSGATGARLKEGAENRSLSITLGRVTAALADM-- 298
| : : : | | | | | | | | : : : : : | : : : | | : : : |
DB 231 DEPTMLSTKSVKISLVDLASGERADSGAKGTRLKEGANINKSLTLTGKVISALAEVON 290
| : : : | | | | | | | | : : : : : | : : : | | : : : |

OY 299 ---SSGQKQKNDLVYRROSVLTWLLKDSIGGNSMTAMTAISPADINEETSTLRYADS 355
| : : : | | | | | | | | : : : : : | : : : | | : : : |
DB 291 CTSKSKKKKKTDFIYRROSVLTWLLRENGNSRTAMVAALSPADINDETSTLTAYADR 350
| : : : | | | | | | | | : : : : : | : : : | | : : : |

OY 356 AKRIKNHAVNEDPNARMIRELKEELAOILRSKLOSOGG-----GGAGS----- 400
| : : : | | | | | | | | : : : : : | : : : | | : : : |
DB 351 AKQICNAVINEDPNAKLVRELKEEYTRILKLRAGLGDIDIDPLIDYSGSGGKYLK 410
| : : : | | | | | | | | : : : : : | : : : | | : : : |

```

```

OY 401 -----SGPYEESYPDTPLEKQ-----IVSIQOPDAPYVK 431
| : : : | | | | | | | | : : : : : | : : : | | : : : |
DB 411 DEONKHHRYLASENQRGNFSTASMSLTS--PSSCSLNSQVGLTSYISQ--BRINST 467
| : : : | | | | | | | | : : : : : | : : : | | : : : |

OY 432 MSKAEIVQDNQSEKLYRDLNQTWEKLAKEEIKERPAALAEELGISIEK--GVGPRH 489
| : : : | | | | | | | | : : : : : | : : : | | : : : |
DB 468 PGGEAIEIRLKSSEKILIELNETWEKLNKTEAINNERALLAEQVAILREGGTLGVYS 527
| : : : | | | | | | | | : : : : : | : : : | | : : : |

OY 490 SKEMPHLVLSDDPLLAECTLYVNIKRGOTRGVNVODTQAEIRLNGSKILKECHTEFENV- 548
| : : : | | | | | | | | : : : : : | : : : | | : : : |
DB 528 PKKTHLVNLNDEKPLMSECLLYIKDGLTRVGOAERRODLYLGAHKEKHECLFRSER 587
| : : : | | | | | | | | : : : : : | : : : | | : : : |

OY 549 ----DNVYTVIENKAAVWNGVRIDKPTRLRSYRIILGDPIHFRFNHPEAREARQO 604
| : : : | | | | | | | | : : : : : | : : : | | : : : |
DB 588 SNTGEVYITLPECSESEYVNGKRAVHPOLRSGNRILINGKHVFRFNPEQABAREK- 646
| : : : | | | | | | | | : : : : : | : : : | | : : : |

OY 605 SLIRSVYNSQLSGPAPRGHDTLSKAGSDADGSRSDPLPHFRKQSDMFTYAREAS 664
| : : : | | | | | | | | : : : : : | : : : | | : : : |
DB 647 -----TPSAETPSPYDWTFAQRELE 668
| : : : | | | | | | | | : : : : : | : : : | | : : : |

OY 665 ALLGDDK-----ISHLTDELDAL-----FDDVQKARAVRGL-----VED 701
| : : : | | | | | | | | : : : : : | : : : | | : : : |
DB 669 K-QGIDMKQEMKRLQEMELLYKKEKEADLLBQORLDYESKLAQLQOVETRLAET 727
| : : : | | | | | | | | : : : : : | : : : | | : : : |

OY 702 NEDSDQSSEFP 712
| : : : | | | | | | | | : : : : : | : : : | | : : : |
DB 728 TEEEEEVEVP 738
| : : : | | | | | | | | : : : : : | : : : | | : : : |

RESULT 5
KF1B_HUMAN STANDARD: PRT; 1816 AA.
AC 06033;
ID 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Kinesin-like protein KIF1B (Kip).
GN KIF1B OR KIAA0591.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OC NCBI_TaxID:9606;
OX 11
RN 1 SEQUENCE FROM N.A. (ISOFORM 1).
RA Bougueret L., Dufaure-Gare I., Grel P.;
RT "DNA encoding a kinesin-like protein (hkip) comprising diallelic
RT markers.";
RN 2 Patent number WO0063375, 26-OCT-2000.
RN 3 SEQUENCE FROM N.A. (ISOFORM 2).
RA Yang H.W., Takita J., Chen Y.Z., Soeda E., Piao H.Y., Hashizume K.,
RA Hayashi Y.;
RT "Cloning of human kif1b gene that maps at 1p36, which is homozygously
RT deleted in neuroblastoma cell line NB1 and mutated in some other
RT neuroblastoma cell lines.";
RN 4 Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.
RN 5 SEQUENCE FROM N.A. (ISOFORM 3).
RA Park M., Shin H., Lee Y.M., Moon E., Choi W., Kim W.;
RT "Identification of the human ortholog of mouse Kif1B, a kinesin
RT superfamily motor protein.";
RN 6 Submitted (Jul-2001) to the EMBL/Genbank/DBJ databases.
RN 7 SEQUENCE OF 479-1816 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RX Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RX Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RN 8 DNA Res. 5:31-39(1996).

```

[5]
 RN SEQUENCE OF 1449-1816 FROM N.A.
 RP Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu K., Nakamura Y., Nagahara K., Masuno Y., Oshima A.,
 RT "NEO human cDNA sequencing project."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 [6]
 RP VARIANT CMT2A LEU-98.
 RX MEDLINE=21885926; PubMed=11389829;
 RA Zhao C., Takita J., Tanaka Y., Setou M., Nakagawa T., Takeda S.,
 RA Hayashi Y., Terada S., Nakata T., Takei Y., Saito M., Tsuji S.,
 RA Hayashi Y., Hirokawa N.;
 RT "Charcot-Marie-Tooth disease type 2A caused by mutation in a
 microtubule motor KIF1B-beta."
 RL Cell 105:587-597(2001).
 CC -1- FUNCTION: MOTOR FOR ANTEROGRADE TRANSPORT OF MITOCHONDRIA. HAS A
 CC MICROTUBULE PLUS END-DIRECTED MOTILITY.
 CC -1- SUBCELLULAR LOCATION: PREDOMINANTLY FOUND ON THE VESICLE- AND
 CC TUBULE-LIKE STRUCTURES WITHIN THE CYTOPLASM (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DISEASE: Defects in KIF1B are a cause of Charcot-Marie-Tooth
 CC disease type 2A (CMT2A). CMT2A is an autosomal dominant neuropathy
 CC characterized by normal or slightly reduced nerve conduction
 CC velocity and axonal loss with little evidence of demyelination or
 CC hypertrophic changes in nerve biopsies.
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104
 CC SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 FHA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AX03604; CAC16629.1; -;
 DR EMBL: AF257176; AAK49332.1; -;
 DR EMBL: AY043362; AAK85155.1; -;
 DR EMBL: AB011163; BAA25517.1; -;
 DR EMBL: AK022977; BAB14341.1; -;
 DR MIM: 605995; -;
 DR MIM: 118210; -;
 DR InterPro: IPR000253; FHA_domain.
 DR InterPro: IPR001849; PH.
 DR InterPro: IPR001752; kinesin.
 DR Pfam: PF00498; FHA.1.
 DR Pfam: PF00225; kinesin.1.
 DR Pfam: PF00169; PH.1.
 DR PRINTS: PR00380; KINESINHEAVY.
 DR SMART: SM00240; FHA.1.
 DR SMART: SM00129; KTSC.1.
 DR SMART: SM00233; PH.1.
 DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE: PS50067; KINESIN_MOTOR_DOMAIN2; 1.
 DR PROSITE: PS50066; FHA_DOMAIN1; 1.
 DR PROSITE: PS50003; PH_DOMAIN; 1.
 KW Motor protein; Microtubules; ATP-binding; Coiled coil; Mitochondrion;
 KW Alternative splicing; Disease mutation; Charcot-Marie-Tooth disease.
 FT DOMAIN 1 361
 FT DOMAIN 365 386
 FT DOMAIN 470 502
 FT DOMAIN 556 612
 FT DOMAIN 668 737
 FT DOMAIN 841 869
 FT DOMAIN 1702 1799
 FT NP_BIND 97 104
 ATP (POTENTIAL).

FT	VARSPPLIC	289	294	MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT	VARSPPLIC	394	434	IDPLIDYSGSSKRYLKDFQNNKHRYLASENORPHFSRA
FT	VARSPPLIC	707	1196	-> T (IN ISOFORM 2 AND ISOFORM 3). YESKLAQLQVETRSIAETTEESEEEVPTQHEPFLA OMAFRRKSHQFSLDLMLGNVYLKEAAVISELKKVQ FOVLLITLTLSPPELLETMEKTHEDPPRTVAAYE ODLKNATHTWLSLEKRLDLKREMYDAGVASSADPS ETTVGSDPYEDRHHMKLVGSSPIFGCVENLARTPSP TESTADSDITELADROODEMDEPDDPDDADGSDCTEBS SDFPSDCHDPFTRSFWEILVGRAPYLSNLPLPYLIHRY AIVSEKGEVGFRLVAQALADEADVDGSGIRQSGTALV SPDNEYFNQDSFSSVMTLRSGLSELEIRIYEGGQSEVLT PPEEIRINDLDSSTLLDGKMEGFESEIENHKLKDEY FTRVYVLOAGSGLIPEVADIFCOFNHRHDEFTSEPLKN NGRGSPLAPYHONIAVEITRESFVDYIKTPYIEVRECH -> ADSDSGDSDSKRSCSESWKILITSLREKLPSPKIQIYK KGLPSSGKRREPIMYQIPQRRRLSKDSKWITISDLQIA VKEICYEVALNDRHSROEIELALIVMKELCMYGRKDPN ERDSWRVARDVMDVGVGDEKIEDVATKSGSDVDLKV HIDKLEDTLOEVKKONNMKDEIKVLRNKLKMEKPLIG SOEOKSPGSHKAKPEYAGACYSSTSENNVSGDGLAKBER VSOLMNGDPFRFRGRRLRMKQEDIRFNLDQDITLOLRQ NVPHRFTIPENRRPREPKSNPHRNSWFGTHIITLEDV IELRIPKDEARKNKEESQEGKGFQDPFGMSQGR SODHIOYSKOHINNOOPOLRMRSNLNNGOPKSTROCS ASAESLNSHSGHPTADQTOFOARKHITHOROSYCNVTGGQ LEGNAATSYKQKDFDQSHCSQFPYTPPRMRQPSAPMLKACR ETTV (IN ISOFORM 3). MISSING (IN ISOFORM 3). Q -> L (IN CMT2A). /FTID-VAR.011515. TNH -> KIN (IN REF. 1). D -> E (IN REF. 1). R -> L (IN REF. 1). VYV -> AVF (IN REF. 1). ILATV -> NLSTE (IN REF. 1). T -> I (IN REF. 1). D -> E (IN REF. 1). I -> N (IN REF. 1). SQ SHQUNCE 1816 AA; 204427 MW; ECAB9647215AC7DB CRC64;

Query Match 40.3%; Score 1633.5; DB 1; Length 1816;
 Best Local Similarity 44.1%; Pred. No. 7.1e-79;
 Matches 349; Conservative 127; Mismatches 178; Indels 137; Gaps 16;

QY	4	GGNKKVYVRRPFRNAREIDGAKCIYVMENQTLITPPGAEKARKSGKTIMGPAFA 63
DB	3	GASVYKVAVRVPFNSRETSKESKCIIOGNSSTINPKNKE-----APKSFS 51
QY	64	FDRSYWGF-DKNAPNVARQEDLPQDLGVPLIDNPFKVCNNCTIFAYGGTGSGKSYSMNGYG 122
DB	52	FDYSYWSHTSPEDPCFASQNRVYNDICKEMLLAFEGYNWCIFAYGGTGAKSTYTMGKO 111
QY	123	KEH---GVIPRICODMFRINELQKKNLTCTVEVSEYIEINERYRDLNPNSTKGNLKVRE 180
DB	112	EESQAGIIPQICELEF-ETNNDCNENEMSYSEVSEWETICERADLDLNPKNKCNLVRD 170
QY	181	HPSTGPYVEDLAKLVRSFOETENLMDGKNKARTVAATNNNETSSRHAVFTLTLOKWH 240
DB	171	HPRLGPYVEDLSKLAVSYTDIADMDAGNKARTVAATNNNETSSRHVYVYITFOTKKH 230
QY	241	DEETKMDTEYAKISVDLNGSERASTSGATGALKGAEINRSLTGLGVIALADM-- 298
DB	231	DNETILLATVSKTSVLDVNGSDRAASTGAGRLKGCALITNLSLTGLGVIALAEVDN 290
QY	299	---SSGRQKNQOLVPRDSVLTWLKDSLGNSMTAMIAISPADINFEETLSTLRVDS 355
DB	291	CTSKSKKKKKTDPLPYDSVLTWLTRENLGNSNSTAVVAALSPADINVDLTSLTRADR 350
QY	356	AKRIKHAENVNEDNAPKRIEELAEOLNSKLOSSGGG-----GGAG----- 399
DB	351	AKQIKCAVINEEDNAKLVRTELKEEVTRLKDLRLAOGIGDIIDPLIDYSGSGSKYLK 410

QY 400 -----GSGPVESSPPDPLEKQ-----IVSIQOPDATYK 431
Db 411 DEONKHRYLLASENQRGHSTASMSLUTSS-PSSCSLSQVGLTSTSIQ-ERINST 467
QY 432 MSAKAIYVQLNLOSELYLDNQTWEKLAETFEIKEREALAEIGISTEK--GVGPPH 489
Db 468 PGGEBAIRLERSEKTIYELNETWEKLRKTEALIMERRALLAENGVAIREDGGTIGVFS 527
QY 490 SKEMPHLVNLSDDPLLAECIYVNIKPGOTRVGNVQDQATQAEIRLNGSKILKECTFEVY- 548
Db 528 PKRPHVLVNMEDPLMSSECLLYITDGITRVQADAEKRODVLVSGAHKEKHCIFRSR 587
QY 549 ----DNVTVIPNEKAAVANGVRIDKPTRLRSYRIILGDEHIFRPNPEBARAEQO 604
Db 588 SNSGCVIYTLPEPCERSEETVYNGKRVSPQVLRSGNRIIMGKHHVRFNPPQARAEREX- 646
QY 605 SLLRHSYVNSQLGSPAPRHRTLSKASDADGDSRSDPLPHFRGKSDWYARREAS 664
Db 647 -----TPSAETPSEPYDWTFAQRELE 668
QY 665 AILGLDQK-----ISHLTDELALFDD-----VOKARAVRGIVED 701
Db 669 K-QGIDMQEMEKRLQEMELLYKKKEKEADLLLEQQRIDYESKLDALQKVTRSLAET 727
QY 702 NEDSDSQSSPP 712
Db 728 TEEEEEVEVP 738

RESULT 6
KPID_RAT STANDARD; PRT: 1097 AA.
AC 035787;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kinesin-like protein KIFID.
GN KIFID.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid-10116.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR;
RA Rogers K.R., Griffin M., Brophy P.J.;
RT "The secretory epithelial cells of the choroid plexus employ a novel
RT kinesin-related protein.";
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 FHA DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AJ000696; CAA04248.1; -.
DR HSSP; P17119; 3KAR.
DR InterPro; IPR000253; FHA_domain.
DR InterPro; IPR001752; kinesin.
DR Pfam; PF00498; FHA.1.
DR Pfam; PF00225; kinesin.1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00240; FHA.1.
DR SMART; SM00129; KISC.1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; FALSE_NEG.
DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.

DR PROSITE; PS50066; FHA_DOMAIN; FALSE_NEG.
KM Motor protein; Microtubules; ATP-binding; Coiled coil.
FT DOMAIN 1 357 KINESIN-MOTOR.
FT DOMAIN 358 380 COILED COIL (POTENTIAL).
FT DOMAIN 437 478 COILED COIL (POTENTIAL).
FT DOMAIN 520 587 FHA.
FT DOMAIN 630 671 COILED COIL (POTENTIAL).
FT DOMAIN 824 868 COILED COIL (POTENTIAL).
FT NP_BIND 96 103 ATP (POTENTIAL).
SQ SEQUENCE 1097 AA; 122333 MW; 8EFA0B1C7579BA5B CRC64;

Query Match 40.1%; Score 1617; DB 1; Length 1097;
Best Local Similarity 45.9%; Pred. No. 7.9e-79;
Matches 357; Conservative 129; Mismatches 184; Indels 108; Gaps 20;

QY 4 GGNIRYVAVRPENAREIDRGAKCIYRMGNCTILTPPGAEEKARKSKITMDGKARA 63
Db 3 GASVKAIVAVRPENAREISQDAKCVVSMGNTSTIINP-----KQSRMFL--KA-S 50
QY 64 FDRSYWSEDK-NAPYARQEDLFODLGVPULDNAEFKYNCCIYAGQGTSGSKYSMMGYG 122
Db 51 FDIYSWSHTSVEDPQFASQOQYVRDQGEMLAHAFEGVVCIFAYQGTAGKSYTMGQ 110
QY 123 K--EHGVIPRICQDMFRINELQDKNLTCYVEVSYLEYINERVRDLNPSKGNLYRE 180
Db 111 EPGQGIYVQLCEDLFESRYN-VNQSALSYSEVSMEYICERVRLDLPKRSGLRYRE 169
QY 181 HPSGTGYVEDLAKLVRSQOEIENLMDENKARVAAATMMNETSSSHAVFTLTLOKH 240
Db 170 HPILGPYVDLSKLAVTSTADIALDMDCGNKARVAAATMMNETSSSHAVFTLVFORSH 229
QY 241 DEFTKMDTEKVAKISLVLDAGSERATSTGATGARLKEGAEINRSJSTLGRVIAALADNS 300
Db 230 DDLTGIDSEKSVKISLVNLAGSERADSSARGMRLEKGANIKSLTTLTKVLSALADLOS 289
QY 301 GKQKNQVLVPYRDSVLTWLLKDSLGNSMTAMIAISPADINFEETLSFLRYADSAKRIT 360
Db 290 -KKRKSDFIPYRDSVLTWLLKKNIGNSRTAMIALSPADINVEETLSFLRYADRTKQIR 348
QY 361 NNAVYNEPNAMIRELKEKELQLRKLSGSGGGGAG-----SGG----- 403
Db 349 CNAVINEPNARLIRELQEVARLRELLMAOGLSASALGGLVEEGSPGGVLPAASSPPA 408
QY 404 PYEESYPPDTPLEKQIVSQPDATYKMSKAEIYVQLNLOSELYLDNQTWEKLAET 463
Db 409 PASPSPP--PINGLEPEPSPSAE-POIGPEAMERLQETEKIYALNETWEKLAET 465
QY 464 EIKEREALAEELGISTIEKGVGYPYKEMPHLVNLSDDPLLAECIYVNIKPGOTRVGNV 523
Db 466 ALRMEREALLAEMGSPGWRIVGVSPPKTPPHLVNLEDPLMSECLYHIKDGVTREVGV 525
QY 524 NODQAEIRLNGSKILKECTFEVY----DNVTVIPNEKAAVANGVRIDKPTRLRS 578
Db 526 DVD---IKLQOFIREQCLFRSLTPQPDGEVAVVLEPEGAEYVNGLVLPVLKSG 581
QY 579 YRIILGDFHIFRPNPEBARAEQOSILRHSVTNSQLSPAGRHDRFLRSAGSDADG 638
Db 582 NRIYVGNKNVFFNPPQARLERER-----GVPPP----- 611
QY 639 SRSDFLPFRGKSDWYARREASAILGLDQKISHLTDELALFDDVOKARAVRGL 698
Db 612 -----PGPPEPYDWMNAQKDWLEQ--GIDIKIE--NEKRLODENQYRKEKEADLL 661
QY 699 VE-----DNEDSDSQSS-----FPVBDKYMSNCTINFSLDTAITMGPPRS 740
Db 662 LEOQRLYADSDSGEDSKRSCESWRLLISLKDLPN-----TVQTVKRCGLPSS 713

RESULT 7
ID U04_CAEEL STANDARD; PRT: 1584 AA.
AC P23678;

Db 224 - ECSEVGLDGENHTRVGLNVLVDLAGSERQAKTGAQGERLKEARKINLSLALGNVISAL 282
QY 296 ADMSSGKOKKQOLVYRDSVLTWMLKDSLGSNSMTAMIAISPADINFEETLSTLRVADS 355
Db 283 VD-----GKSTHPIYRDSKTRLLDQSLGNSNATVAVANGPASYNVEETITLIRYANR 336
QY 356 AKRIKNAHAVNEDPNAMRIELKEELAQLRSKLOSSGG-----GGAGSGSGPV 405
Db 337 AKNIKRPVNEDEPKDALREFOEIARLKAQLEKRSIGRRKRRKREGGSGGSGGEE 396
QY 406 EESTPDPPLPEKQIVSTQOPDATYVKMSKATIVG-----LQNSEKLYRLDNQWEEK 458
Db 397 EEEEGE-----GEEEGDDKDDYWRQOEKLEIKRAIYEDHSLVAEEKM-----RLKEK 446
QY 459 LAKTEEIKKEREALEELGISIE-----KGFVGPY-----HSKEMPHLVNLSDDPLAECLV 510
Db 447 EKKMEDLRREKDA-EMIGAKIKAMESKLVGKNIVDHTNEOKILQKQCEIAEQ----- 502
QY 511 YNIRPGOTRVGNVODTQAEIRLNGSKILKEHCETFEVNDVNTIVPNEKAAVNGVRID 570
Db 503 -----KREREIOQ--QMESRDEETLEKE--TYSSIQOEVDI-----KTK 539
QY 571 KPTRLRSGRYIILGDHFIRFNHPREARABERQOS-----LIRHSVNSQLGSPAPGRH 624
Db 540 KLKLFSLQAVKAEIHDQEEHTEKEROELQOTONELTREIKLHLIENFI--PLEEK- 596
QY 625 DRTLKAGSDADGDSRSDPLPHFRGKD-----SDMEFYAR-----R 660
Db 597 SKINNRAPFDEEDHMKLHPITRLNQOMMKRPYSAVGKRPPLSHARMSMIRPEARXR 656
QY 661 EASAIIGLDOKISHLVDDELDALEFDVQKARAVRGLVEDNESDSOSSF 711
Db 657 AENITLLELDMPSTRTRDYEGPATAIPKVO--AALDALQDEDELDVQDASSF 705

RESULT 11
KF3B_MOUSE STANDARD; PRT: 747 AA.
ID KF3B_MOUSE
AC Q61771;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE kinesin-like protein kif3b (microtubule plus end-directed kinesin motor 3b).
GN KIF3B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ICR; TISSUE-Brain;
RX MEDLINE=96032268; PUBMED=7559760;
RA Yamazaki H., Nakata T., Okada Y., Hirokawa N.;
RT "Kif3b/B: a heterodimeric kinesin superfamily protein that works as a microtubule plus end-directed motor for membrane organelle transport.";
RL J. Cell Biol. 130:1387-1399(1995).
CC -1- FUNCTION: MICROTUBULE-BASED ANTEROGRADE TRANSLATOR FOR MEMBRANOUS ORGANELLES. PLUS END-DIRECTED MICROTUBULE SLIDING ACTIVITY IN VITRO.
CC -1- SUBUNIT: HETERODIMER OF KIF3A AND KIF3B.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN II SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-slb.ch/announce/> or send an email to license@slb.ch).
CC -----

DR EMBL; D26077; BAA05070.1; -.
DR HSSP; P17119; 3KAR.
DR MGD; MG1:107688; KIF3b.
DR InterPro; IPR001752; kinesin.
DR Pfam; PF00225; kinesin_1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; Kisc_1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Microtubules; ATP-binding; Coiled coil; Neutone.
FT DOMAIN 1 345
FT DOMAIN 346 579
FT DOMAIN 580 747
FT NP_BIND 96 103
FT DOMAIN 386 393
FT DOMAIN 394 405
FT DOMAIN 723 730
FT POLY-SER.
SQ SEQUENCE 747 AA; 85288 MW; FA369A4190EC8B47 CRC64;

Query Match 21.0%; Score 844.5; DB 1; Length 747;
Best Local Similarity 32.9%; Pred. No. 6,66-38;
Matches 256; Conservative 116; Mismatches 253; Indels 153; Gaps 27;
6 NIKVYVRVPFNARE---IDGAKCIVRMGEGNQTILTPPGAEEKARKSGRTINDGPKA 61
Db SVRVVVRCPMNGKKAASYDKVDVVKL--GQSVAKNPKG-----TSHEMPRT 56
9
62 FAFDSTY-WSPDKNAPNARQEDLFQDGLVPLIDNAFGYNNCTFAYCOTGSGKSYSMG 120
Db FTFDAVYWMN-----AKQFELYDEFRLVYDSVLOGFNQITFAYGOTGKTYTMEG 108
57
121 Y-----GKEGVIPRLICODMFRINELQDKNLTCTVEGVYLETYNVRVDLNPSTGNLK 177
Db VGRDEKRGVLPNSFDHFTHSRSQNOYL--VRASYLEYQEIIDLSDKDTKRL 165
109
178 VREHPSGPVYEDLAKLVYRSFOEITENLMDGKNKARTVAATNNETSSRAVFTLTITQ 237
Db LKERPDGTGVYVVDLSSFTKSVKEIEHVWANGNQNSRGATMNMHSSRAHIFVITI-- 223
166
238 KWHDETQMORE---KVAKISLYDLAGSERATSPGATCARLKEGAEINRSLTGRVTA 294
Db 224 -ECSEVGLDGENHTRVGLNVLVDLAGSERQAKTGAQGERLKEARKINLSLALGNVISAL 281
295 LADSSGKOKKQOLVYRDSVLTWMLKDSLGSNSMTAMIAISPADINFEETLSTLRVAD 354
Db 282 LVD-----GKSTHPIYRDSKTRLLDQSLGNSNATVAVANGPASYNVEETITLIRYAN 335
355 SAKRIKNAHAVNEDPNAMRIELKEELAQLRSKLOSSGG-----GGAGSGSGPV 404
Db 336 RAKNIKRPVNEDEPKDALREFOEIARLKAQLEKRSIGRRKRRKREGGSGGSGGEE 395
405 VE-----ESTPDPPLPEKQIVSTQOPDATYVKMSKATIVG-----LQNSEKLYRLDNQWEEK 458
Db 396 EEEEGE-----GEEEGDDKDDYWRQOEKLEIKRA--TVEDHSLVAEEKM-----RLKEK 447
459 LAKTEEIKKEREALEELGISIE-----KGFVGPY-----HSKEMPHLVNLSDDPLAECLV 511
Db 448 KKMEDLRREKDA-EMIGAKIKAMESKLVGKNIVDHTNEOKILQKQCEIAEQ----- 502
512 NIKPGOTRVGNVODTQAEIRLNGSKILKEHCETFEVNDVNTIVPNEKAAVNGVRID 571
Db 503 -----KREREIOQ--QMESRDEETLEKE--TYSSIQOEVDI-----KTK 540
572 PTRLRSGRYIILGDHFIRFNHPREARABERQOS-----LIRHSVNSQLGSPAPGRH 624
Db 541 LKLFSLQAVKAEIHDQEEHTEKEROELQOTONELTREIKLHLIENFI--PLEEK- 596
618 -----SPADGRDRTLKAGSDADGDSRSDS-----PLPHRGKDS 653
Db 601 NRSFDDDEDHMKLHPITRLNQOMMKRPYSAVGKRPPLSHARMSMIRPEARXR- 659
654 DMFYARREASAIIGLDOKISHLVDDELDALEFDVQKARAVRGLVEDNESDSOSSF 711

Db 660 -----IMLEEDMPRTTDEGPAISPKVO--AALDAALQDEDEIQDASSF 705

RESULT 12

K12L_STRPU STANDARD; PRT; 742 AA.

ID K12L_STRPU STANDARD; PRT; 742 AA.

AC P46871; 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Kinesin-II 95 kDa subunit (KRP-85/95 95 kDa subunit).

GN KRP95.

OS Strongylocentrotus purpuratus (Purple sea urchin).

OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.

OC NCBI_TaxId=7668;

OX NCBI_TaxId=7668;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Egg;

RX MEDLINE=94050179; PubMed=8232586;

RA Cole D.G., Chinn S.W., Medaman K.P., Hall K., Vuong T., Scholey J.W.;

RT "Novel heterotrimeric kinesin-related protein purified from sea urchin eggs.";

RT Nature 366:268-270(1993).

RL -1 SUBUNIT: PRIMER OF A 115 kDa SUBUNIT AND TWO KINESIN-LIKE SUBUNITS OF 95 AND 85 KDA.

CC -1 PTM: THE N-TERMINUS IS BLOCKED.

CC -1 SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN II SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U00996; AAA87393.1; -.

DR HSSP: P17119; 3KAR.

DR InterPro: IPR001752; kinesin.

DR Pfam: PF00225; kinesin; 1.

DR PRINTS: PR00380; KINESINHEAVY.

DR SMART: SM00129; KISC: 1.

DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.

DR PROSITE: PS50067; KINESIN_MOTOR_DOMAIN2; 1.

KW Motor protein; Microtubules; ATP-binding; Coiled coil.

FT DOMAIN 1 337 KINESIN_MOTOR (BY SIMILARITY).

FT DOMAIN 338 613 COILED COIL (BY SIMILARITY).

FT DOMAIN 614 742 GLOBULAR (BY SIMILARITY).

FT NP_BIND 95 102 ATP (POTENTIAL).

FT SEQUENCE 742 AA; 84202 MW; 47640A367BA77B5 CRC64;

Query Match 20.5%; Score 827.5; DB 1; Length 742;

Best local similarity 31.5%; Pred. No. 5, 2e-37;

Matches 250; Conservative 122; Mismatches 286; Indels 135; Gaps 25;

QY 181 HPSTGYVEDLAKLVYRSFQEIENLMDENKARTVAATNNETSRSRHAVFTLLTQKWH 240

Db 168 RPTGVYVVDLSFVTRKSVKIEHVMVTGVNNRNSVGSNTNMNHSRSHAFITTI-----E 223

QY 241 DEETKMDTE---KVAKISLVDLGSEKATSTGATGARKGAGATNSLSLGVATLAAD 297

Db 224 CSELGYDGENHIEVGLKNTLMDLAGSEKQATGATGRLKATKINISLSLGVATLAAD 283

QY 298 MSSGKQKNOVLVPRSVLWKLKDSLGSNSMTAMIAISPADINFEETLSTRYADSAK 357

Db 284 -----GKSHIYRDSKLTLLRLQDSLGNKATVYMANMGPASYNDETITTLRYANRAK 337

QY 358 RIRNHAIVNEDPNARMIRELKEELQSLKQSSG--GGGGAGSGGAPVEEYPPPTPLE 416

Db 338 NIKNKPINPDMDALREFPEEISRLKQALDKKPSDGRKKKKRPPGQGDDEDE 397

QY 417 KQIVSTQPPATYKKMSKATVEQLNSELVYDNLQNTWEKLAKEEI-----H 466

Db 398 TEEEGDEMEDEMYKESQKLEE--EKEMINQSKIAEKOKLSEVQKQGEIKKEH 454

QY 467 KEREALAEELGISLEKGFVGPY-----HSKEMPHLVNLSDDPLAECLVYNIKPGQTRVGN 522

Db 455 QQKEMLEKTKAKESKILVGGKSTVDHTNEQKRIE--EQRLLLAE-----EKN 501

QY 523 VNQDTQAEIRLNGSKILKECHTEFENDVNTIVPNEKAAYVNGVRIDKPTRLRSGRIT 582

Db 502 RERDMERKLEQDDKVEIGETFSLSQEEV-----VTKRLKLKLFKALQSY 548

QY 583 LGDFHIFRHNPEEAR--AEQEOSILRHSTNSQLSGPARGHRT--LSKASDADGDS 639

Db 549 KSDIQDLQDDHAREROLEOTQNELRELKRVADNFIPEVERKITTVAVDEETEE 608

QY 640 RSDPLPFRGKD-----SDWFYAR-----REASAIILGLDOKIS 674

Db 609 WLTPPLAKAGSPQMAKRPVSAYGNRRPAD--YARAAQMGCPKRAKNILISVLDMP 666

QY 675 HLT--DDELDALEDDYOKARVRGLVEDNEDSDQSFSFY-----BDKYMSNGTID 724

Db 667 NRTTRDVEGSAVAPRVQ--AALDAALQDEDDDLLEVQPVFAKTKLKKDKYRSK---- 719

QY 725 NFSLDATITMPTGT 737

Db 720 ---HKAVAKPGS 728

RESULT 13

FL10_CHLRE STANDARD; PRT; 786 AA.

ID FL10_CHLRE STANDARD; PRT; 786 AA.

AC P46869; 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Kinesin-like protein FLA10 (Khp1 protein).

GN FLA10.

OS Chlamydomonas reinhardtii.

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

OC NCBI_TaxId=3055;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=137;

RX MEDLINE=94299638; PubMed=8027176;

RA Walther Z., Vashishtha M., Hall J.L.;

RT "The Chlamydomonas FLA10 gene encodes a novel kinesin-homologous protein.";

RT J. Cell Biol. 126:175-188(1994).

RL -1 FUNCTION: PROBABLY INVOLVED IN FLAGELLAR ASSEMBLY AND MAINTENANCE.

CC -1 TISSUE SPECIFICITY: FLAGELLAR AXONEME.

CC -1 SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN II SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: L33697; AAA21738.1; .
DR HSSP: P17119; 3KAR.
DR InterPro: IPR001752; Kinesin.
DR Pfam: PF00225; Kinesin: 1.
DR PRINTS: PR00380; KINESINHEAVY.
DR SMART: SM00129; KISC: 1.
DR PROSITE: PS00411; KINESIN.MOTOR.DOMAIN1; 1.
DR PROSITE: PS00411; KINESIN.MOTOR.DOMAIN2; 1.
KW Motor protein; Microtubules; ATP-binding; Coiled coil.
FT DOMAIN 1 358 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 367 687 COILED COIL (POTENTIAL).
FT DOMAIN 688 786 GLOBULAR (POTENTIAL).
FT NP_BIND 97 104 ATP (POTENTIAL).
FT DOMAIN 388 391 POLY-GLY.
FT DOMAIN 705 714 POLY-GLY.
FT DOMAIN 756 759 POLY-ASP.
SQ SEQUENCE 786 AA; 86671 MM; F90969203EB79F1B CRC64;

Query Match 20.5%; Score 825.5; DB 1; Length 786;
Best Local Similarity 39.3%; Pred. No. 7.2e-37;
Matches 204; Conservative 80; Mismatches 168; Indels 67; Gaps 14;

OY 2 SGGG--NIKVVVRVPPFNAREIDRGAACIVRNE---GNOTITTPPGAEERKARKSGKTIM 56
DB 4 AGGGSESVKVVVRCPLNGKEKADRSRLVDMVDVAGOVKVPADASEP----- 54
OY 57 DPKAFAPDRSY-WSEFDKNAIPYARQEDLPDLGVPLDPAKFGYNNCFAYGOTSGSKS 115
DB 55 --KAFETPDQYVDWV-----CQRDVFDITARPDLIDSCIEGNTNGTIFAGGTGTGKS 104
OY 116 YSMGYSYKE--HGVIPIRICODMFRINELQDKMLTCTVEVSYLEYINERVRDLLNPST 172
DB 105 HTMEGKDEPPELRGLIPMFERYVFELIARDSGTKEF--LVRSSELEYINEVRDLLGKDH 162
OY 173 KGNLKVREHPSRGPVVEDLAKLVRSFOEIENTLMBEGNKARKYVAATNNMETSRSRAVVT 232
DB 163 SKKMKELKSPDGKGVYVKDLSQFVCKNYEEMKRVLLAGKKNROYGATLNMQDSRSISFT 222
OY 233 LTL-----TQMHDEETKMDTEKVAKISLVLAGSEBATSNGAGTGARLKEGAEIN 282
DB 223 ITIEIEKLESAAKPKPAKKDSNHYRGKLNVLVDLAGSEKQDTGATGDKLKEBGIKIN 282
OY 283 RSLSTLGRVIALADMSGCKOKNQLVPYRDSVLTWMLKDSLGNSMPTAMIAISPADIN 342
DB 283 LSLTALGNVIALVDGKSGH-----IPYRDSKLTFRLLQDSLGNTKTVYVANITGPADMN 336
OY 343 FEETISTLRYSADAKRIKIHAVNEDPNARMTRELKEELAQRSKLOSGGGGGG---A 398
DB 337 YDETSTLTATYARAKNIQKPKINEDPKDAMLRQFOEIEIKLKEQLAAAGAGGGGTTPM 396
OY 399 GSGSGPVE-----ESYPPD-----TPEKQIVSYIQQPDATVYKKSKAEIYEQLNQ 443
DB 397 SGGSGPPTOKIYERTEVDDPIDAIAQMAELDAKKMSISTEALDKAREEAALAK-KQ 455
OY 444 SEKLVRDLNQWTEKLAETETIKEREALAEELGISIEK 482
DB 456 LQAIIDQCKTEAOKKARDAALKQAEEARALAG-AIEK 493

RESULT 14
KINL_LEICH STANDARD; PRT; 955 AA.
AC P46865;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Kinesin-like protein K39 (Fragment).
GN KIN.
OS Leishmania chagasi.
OC Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5669;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HOM/BR/82 / ISOLATE BA-2;
RX MEDLINE=93133867; PubMed=8421715;
RA Burns J.M. Jr., Shreffler W.G., Benson D.R., Chailb H.W., Badaro R.,
RA Reed S.G.;
RT "Molecular characterization of a kinesin-related antigen of
RT Leishmania chagasi that detects specific antibody in African and
RT American visceral leishmaniasis."
RL Proc. Natl. Acad. Sci. U.S.A. 90:775-779(1993).
CC -I- DEVELOPMENTAL STAGE: PREDOMINANT IN AMASTIGOTES.
CC -I- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: L07879; AAA29254.1; .
DR HSSP: P17119; 3KAR.
DR InterPro: IPR001752; Kinesin.
DR Pfam: PF00225; Kinesin: 1.
DR SMART: SM00129; KISC: 1.
DR PROSITE: PS00411; KINESIN.MOTOR.DOMAIN1; 1.
DR PROSITE: PS00411; KINESIN.MOTOR.DOMAIN2; 1.
KW Motor protein; Microtubules; ATP-binding; Coiled coil; Repeat.
FT DOMAIN 1 399 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 426 >955 COILED COIL (POTENTIAL).
FT NP_BIND 122 129 ATP (POTENTIAL).
FT DOMAIN 704 >955 7 x 39 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 704 742 1.
FT REPEAT 743 781 2.
FT REPEAT 782 820 3.
FT REPEAT 821 859 4.
FT REPEAT 860 898 5.
FT REPEAT 899 937 6.
FT REPEAT 938 >955 7 (PARTIAL).
FT NON_TER 955
SQ SEQUENCE 955 AA; 106168 MM; 8CA76815BE84C6B9 CRC64;

Query Match 20.2%; Score 814; DB 1; Length 955;
Best Local Similarity 28.1%; Pred. No. 3.9e-36;
Matches 253; Conservative 125; Mismatches 292; Indels 230; Gaps 24;

OY 7 IYVVVRVPPFNARE--IDRGACIVRMENQITLPP-----PGAERKARKSGKTIND 57
DB 13 VYVSAYVRPLNRENNAPRGFTKYVAKQAAVAVTVKVLGSGNSGAASMGARRAVD 72
OY 58 GKAPAFDYSYSP---DKNAPRYARQEDLPDLGVPLDPAKFGYNNCFAYGOTSGSK 114
DB 73 ----TQFDHVFWSVETPDACATPATQADVFRITGTPLOVHAFFDGNLSCLFAYGOTSGSK 128
OY 115 SYSMMG-----YGRKHGVIPIRICODMFRINELQDKMLTCTVEVSYLEYINERVRDL 168
DB 129 YTTMGADVASLSEBNGTTPRICLEIFAPKASVYEAQSGSRMTVELGYEVINERKSDLL 188
OY 169 NPSTKG-----NLKVRHPSRGPVVEDLAKLVRSFOEIENTLMBEGNKARKYVAATNN 220
DB 189 GRRKGVKGGGEVYVDVVEHPSRGVFLGQRVLEVGSLDDVVRILTEIGNVYRHTASTK 248
OY 221 NETSSSRHAVVFTLT-----TQMHDEETKMDTEKVAKISLVLAGSEBATSNGAGTGARL 276
DB 249 NDRSSRHAIIMLLKEERTMTTKSGETIRTAGKSSRMNLVLDLAGSERVAQSOVEGOQPK 308

[illegible]

Db 569 HOREIEGLELNTROLSRELRLOMLIIDNFIPODYOEMIENYVHMNEDIGEWJLKCVAITG 628
OY 565 NGVRIDKPT---RLRSGYRIIGDFHIFRFNHPPEARAROEOSILRHSTNSQLGSPAP 621
Db 629 NNMKQOTVPDPKKEHDPPEVDLS--HY-LAYTEESL--ROSLMKLERPRITSKGKARPRT 683
OY 622 GRHDRT 627
Db 684 GRKRS 689

Search completed: July 25, 2002, 05:46:59
Job time: 448 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2002, 05:44:26 ; Search time 70.23 seconds
(without alignments)
1931.198 Million cell updates/sec

Title: US-09-235-416-1

Perfect score: 4030
Sequence: 1 MSGGNIKVVVRVPFNARE.....ELRQOAOAMEALTKAQEF 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_prodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1635	40.6	1773	5	09V776
2	1630.5	40.5	1671	5	09V776
3	1590.5	39.5	1581	5	018778
4	1525.5	37.9	2205	5	09NGO2
5	1468	36.4	1749	11	09ECW7
6	1465	36.4	1749	4	09H193
7	1465	36.4	1757	4	09H194
8	1463.5	36.3	1770	4	09H198
9	1463.5	36.3	1805	4	09H198
10	1412	33.0	1826	4	09H198
11	1396.5	34.7	1921	5	001349
12	1396.5	34.7	1921	5	09V7C9
13	1284.5	31.9	1595	5	020888
14	1174.5	29.1	1648	4	015058
15	1147	28.5	928	5	009997
16	1128	28.0	1174	5	09VB25

17	1076.5	26.7	1499	13	09DDA6
18	1035	25.7	1013	5	0960F4
19	1035	25.7	1121	5	09V1P4
20	1035	25.7	1121	5	016866
21	1034	25.7	1121	5	018390
22	1019.5	25.3	676	3	09C2M3
23	1004	24.9	429	4	09BOM5
24	978.5	24.3	1238	4	09HC12
25	847.5	21.0	752	4	09NKN9
26	807.5	20.0	1038	11	099PW8
27	799	19.8	744	13	093478
28	796	19.8	2756	10	09LJ50
29	783	19.4	699	13	098711
30	781.5	19.4	677	5	09VRK9
31	779.5	19.3	976	5	0969C2
32	775	19.2	1229	10	09LXV6
33	769.5	19.1	672	5	096V93
34	766	18.9	782	5	096V92
35	763	18.9	1099	5	09U142
36	761	18.9	782	5	019633
37	753.5	18.7	646	5	0965T6
38	750.5	18.6	784	5	0961H5
39	748	18.6	735	5	09U0D5
40	744.5	18.5	644	5	096V90
41	743.5	18.4	697	5	09U921
42	735.5	18.3	927	3	09HES9
43	735.5	18.3	1128	5	09U179
44	728.5	18.1	2158	10	09LUM5
45	722	17.9	929	3	P78718

ALIGNMENTS

RESULT	ID	ALIGNMENTS
1	09V776	PRELIMINARY; PRT; 1773 AA.
AC	09V776	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)
DE	CG8566	PROTEIN.
GN	UNC-104	OR CG8566.
OS	Drosophila melanogaster	(Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RC	SEQUENCE FROM N.A.	
RC	STRAIN-BERKELEY;	
RX	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zheng Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,	
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,	
RA	Abdel J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,	
RA	Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhargava D., Bolshakov S.,	
RA	Borova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,	
RA	Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Foster K., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,	
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,	
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,	
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,	


```
Db 55 YSYMSHHDDHDDFSTQSNVYVDIGGEMLQHSFDGYNVCIFAQGYGAKSSTMMRGQEQ 114
Qy 125 -HGVIPIRCODMFRRINELQDKNLCTVEVSYLTYENRERVADLNLPSTKGLKREHPS 183
Db 115 QECITPMLCKDLFRIOPTERD-DLKYSEVSYMEIYCERVDLLNPKNGKLARKREHPL 173
Qy 184 TGPVVEDLAKLVRSFOEIEMLDEGNKARTVAATNMNETSSRSNAVFTLTJQRMDEE 243
Db 174 LGPVYEDLSKLAVTDYODIHLIDEGNKARFVAATNMNETSSRSNAVFTLTJQRMDEE 233
Qy 244 TKMDTEKAKISLVDLASERATSGATGARLKEGAEINRSLSITGRIYALADMSGKQ 303
Db 234 TNLTEKYSKISLVDLASERADSTGAGTRLKEGAMNKSITLTGKYISALAEVYSASKK 293
Qy 304 ---KKNLVPRDSVLTTLKDSLGNSMTAMIAISPADIFEEETLSLRADSAKRK 360
Db 294 KNTKKADIPPRDSALTYLLRENLOGNSKTAMIAISPADINDETSLTRADAKOIY 353
Qy 361 NHAUVNEDPNARMIRELKEELAQLRSKLQSSGGGGAGSGGPGVEESYPPDTPLEKOIY 420
Db 354 CKAUVNEDANAKLIRELKEEIOKLRDLKAEG-----I 386
Qy 421 SIQOPD-----ATYKKMSKAE-----IVEQLNOSKLYRDLNQTWEKLTATEERH 466
Db 387 EVQEDDELTKSTVLSKSPKSRNRNGSTTEMAVVDLQASEKILAEINTEWEKLTATEER 446
Qy 467 KERENALEELGISI-EKGF-VGPRYSKEMPHLVNLSDDPLAECLVYNIKPGQTRGVNV 524
Db 447 VQREAVFAMKGVAVKEDDITGVSPKTPHVLVNLNEDPNLSECLLYIKELTTLRGTHE 506
Qy 525 QDTQAEIRLNSKSLIKECTEENDVNTVTIYPNEKAAVNVCGRIDKPTLRLSGRIILG 584
Db 507 ANVPDDIOLSSHLKECTEENKSTVTLPLPKDAIIVYNGRKLYEEBVLKTGSRIYLG 566
Qy 585 DEHIFREHPEAREBERQESILRHSVTNSQLGSPAPGRHDTLSKASDADGDSRSP 644
Db 567 KNHVFRFTNPQARE-----LRDKT-----ETENEAEVEV--KTDI- 601
Qy 645 LPHFGKDSDFYARRENAASAILGD-----QKISHLTDD-----ELDALFDDVOK-- 690
Db 602 -----QOVNMFQACELLEK-QGIDLKAEMKKRLDNLEBOYKREKRLAQDOFEOGRITY 654
Qy 691 -AR-AVRGLVEDNEEDSOSSPFYRKDYNSNGTIDNLSLDTAITMP----- 735
Db 655 EARIDALOKOVEOSMTSMYSYSPEDFHOEDVYTTNPRESCTAREAGIAAANAFRW 714
Qy 736 -----GTPRSDDDDALFPFGDKSKODASNDVYELRQOQAG 772
Db 715 RYHOFTSLRDDLMGNATFL-----KEANAISVELLKKRVQFO 750

RESULT 3
ID 018778 PRELIMINARY; PRT: 1581 AA.
AC 018778;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HYPOTHEICAL 179.4 KDA PROTEIN.
GN C52E12.2
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peioderinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA MEDLINE-99069613; PubMed-9851916;
RA None;
RT *Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.
RL Science 282:2012-2018(1998).
```

```
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Du 2.;
RT "The sequence of C. elegans cosmid C52E12."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U50135; AA93453.2; -.
KW Hypothetical protein.
SQ
SEQUENCE 1581 AA; 179408 MW; 61CA9CA0B322CCCC CRC64;
```

Query Match 39.5%; Score 1590.5; DB 5; Length 1581;
Best Local Similarity 44.7%; Pred. No. 1.2e-83;
Matches 356; Conservative 139; Mismatches 213; Indels 89; Gaps 19;

```
Qy 6 NIKVYVRPPEARELIDGAKCIYMEGNOTLTTPPGAEAKRSKGTINDGPKAFAD 65
Db 3 SYKVAVRVPFQOREISNTSKCIVLVNGMTTIT-----NGHSINRENSFND 50
Qy 66 RSYWSPDKNAPYARQEDLFQDLGVPLDNMAFKYNNCIFAYGOTGSGKSYSMGYK-- 123
Db 51 HSYWSEFARNDFPFIQKQYVELGYEMLEHAEGYVNCIFAQGYGSGKSYTMGKANDP 110
Qy 124 -EHGYIPIRCODMFRRINELQDKNLCTVEVSYLTYENRERVADLNLPSTKGLKREH 182
Db 111 DEMGIIPRLCNDLFARIDN-NNDKDOYSEVSYMEIYCERVDLLNPKNGKLARKREH 169
Qy 183 STGPVYEDLAKLVRSFOEIEMLDEGNKARTVAATNMNETSSRSNAVFTLTJQRMDE 242
Db 170 ILGPIYVDLTKRACVSYNDICHLMDGKAKRTVAATNMNETSSRSNAVFTLTJQRMDE 229
Qy 243 ETKMDTEKAKISLVDLASERATSGATGARLKEGAEINRSLSITGRIYALADMSGK 302
Db 230 DSNLDTKHSKISLVDLASERANSTGAGRLKEGANINKSLTTLGLVYSKLAESTBK 289
Qy 303 OKKNO-LVPIYRSVLTWLLKDSLGNSMTAMIAISPADINFEETLSLRADSAKRK 361
Db 290 KSNNGGVIPRDSVLTWLLRENIGNSKTAMIAISPADINDETSLTRYADRAKOIYC 349
Qy 362 HAVVNEDPNARMIRELKEELAQLRSKLQSSGGGGAGSGGPGVEESYPPDTPLEKOIYS 421
Db 350 QAVVNEDPNAKLIRELNEVITLRHLKDG-----IDVTD 385
Qy 422 IQQPDATYKKMSKA-----EIVEQLNOSKLYRDLNQTWEKLTATEERHKEERALEEL 476
Db 386 VQETPGKHKKGPRLPAHVHEQLEKLOESEKLMALIGKTMEOQLIHTEELRQREBELRNM 445
Qy 477 GIS-IEKG-FVGPIYSKEMPHLVNLSDDPLAECLVYNIKPGQTRGVNVNTOAEIRLN 534
Db 446 GLACAEDEGTGLGVSPKPKLPHVLVNLNEDPLMSECLITYLKEVTSVGRVAAHRRDILLS 505
Qy 535 GSKIRKEHCTEENNVNTVTIYPNEKAAVNVCGRIDKPTLRLSGRIILGDEHIFREHNP 594
Db 506 GAILLELHCEPINEGANTVLTIMKPNASCTINKOVTTTPVLTGSRVILGEGHHVFRINDP 565
Qy 595 EBARAEEOGOSILRHSVTNSQLGSPAPGRHDTLSKAGSDADGDSRSDPLPHFGKDS 654
Db 566 QBAROSRHILAIARQPIDMKYA-----QOELLDKOGIDLKADMEK-----KMLEMESQ 614
Qy 655 WFIARREASALIGDOKSHITDELDELAPDDVOKARAVR-----GLVEDNED--SD 706
Db 615 Y---RREKVE---LEQKYHOT-REYESMIENLOKOVLAOSYISGGGSIIEGGRMLTS 666
Qy 707 SOSSEFPVRDKYNSNGTIDNFSLDTAITMP-----GTPRSDDDDALFPFGDKSKODASND 762
Db 667 SLIEPPEELKWTSDQ--KRVYLKAALIKRYYHQTYSVRDDLMGNATFV-----KEANAIS 718
```


[illegible]

DR Pfam, PF00225, kinesin. 1.
DR PRINTS, PRO0380, KINESINHEAVY.
DR SMART, SM00240, FHA. 1.
DR SMART, SM00129, KISC. 1.
DR PROSITE, PS50067, KINESIN_MOTOR_DOMAIN2. 1.
DR ATP-binding, Coiled coil, Microtubules, Motor protein.
SQ SEQUENCE 1749 AA: 196070 MW: F4C1C2E8435D578D CRC64:

Query Match	36.4%;	Score 1465;	DB 4;	Length 1749;
Best Local Similarity	43.6%;	Pred. No. 2.6e-76;		
Matches 343;	Conservative 119;	Mismatches 205;	Indels 120;	Gaps 19;

Db	6	VAVANRVRPMNRRELELWTKCYVEHEGQYVLAHPPSMTKQGERK-----PKRYFAFDY	59
Qy	67	SYWSDDK-NAPNYARQEDLFODLGVPLLDNAFKYNNCIPAYGOTSGSKSYMMGCKYEH	125
Db	60	CWMSADESNTTYYAQOEYVFKCJGGLIEKAFQGNACIFAYGOTSGSKSYMMGHEQUL	119
Qy	126	GVYPICDMPRIEILKDKMLCTVYSEVSYLIEHEVBDLLAN-SFKGNLKYAREHST	184
Db	120	GLIPPLCALCFRRIS-LEONOSQTFKVEVSYMEILNEYVRDLDLADKGRSOSLKRAREHYL	178
Qy	185	GPYVEDLAKLVYRSPQEIENLMEDEGNKARVYATNMNETSSKSHAVFLTLTQKHDEET	244
Db	179	GVYGDLOSLAVTSPEDIESLMSBEGNKSRUYAATNMNESSSHAVFNIIITQTLVDLOS	238
Qy	245	KMDTEKRVAKISLVDLAGEBRATSTATGARLKEGAEIIRSLSTGLRVYALALDMSGKOK	304
Db	239	GNSGCEKSVSVSLVDLAGSERVSKTGAABERLKEGSGINIKSLTTLGLVYSLADQAAKG-G	297
Qy	305	KQOLPVPYRPSVLTWLLKXSLGNSMTYALAIISPADIINEEPLSTLRYADSKRRKKNAY	364
Db	298	KSKFPYPRDSVLTLLKDNLGNOSGTSMTATISPADNYEETLSTLRADRRKRRVYNAY	357
Qy	365	VNEDPNARMIRELKEELALQRLSKLOSSGGGGGAGSGGVPYEESSYPDDTPLEKQIVSIQO	424
Db	358	VNEDPNAKVIRRELREVEKRLREQLSOA-----	384
Qy	425	PDATYKKMSKAIYVQLOMSEKLYLDLQWTEPEKLATEEIHKEKREALDEELGISIEKG	484
Db	385	-----EAMKAPELKKELESEKELIELVYWEKIRKTEELAEOERQOLESMGSLSEMS-	438
Qy	485	VGPYHSKEMPHLYVNDSDPLAECLVYVNIKPGQTRVGVVNDQTOAEIRLANSKILKECHT	544
Db	439	-GIRKGDCKCYLVNANADPALNELLYLTK-DHTRVY---ADTSQDIDLFEGIGLOPQICE	493
Qy	545	FENV-DNVYTYVPEKKAIVMNGVRIIDKPTRLRSGYRIILDDFHIFRNPHEARAEQOE	603
Db	494	IDIASDGQVTLTPKNNARSCVNGTLVYCSTQTOAHMDRILMGNNNHFRRLNLPKRKRBDLX	553
Qy	604	QSLHRSVYTNOSLGSFAPGRHDRTLKSGSDGDSRSRSDPLPHRGKSDMVEYARRBA	663
Db	554	D-----FEKETGPR-----EHD-----LDAASASSE-----PDYTERFQMEVY	588
Qy	664	SAILGLDKRISHLTDDELDALEFDVQKARAARGLVEDNE-----DSDSOSFPPV	713
Db	589	MKTLMSNDPQVQVW-VQYLEKQYLEEKRSALJEORLMTYERRELEQLHQQLSPDPOQSSCPD	647
Qy	714	RKYVNSNGTIDNFSLDTATYTMGPTRRSDDDGDALEFGGKSKQOASNDVDEELNQOQAOQ	773
Db	648	RLAYS-----QTA-----OQKTYQAAEERD-ELFPLQSLAKL	678
Qy	774	EEALKTA 780	
Db	679	REQOLVKA 685	
RESULT	7		
ID	09H194	PRELIMINARY: PRT: 1757 AA.	

AC Q9H194;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE KINESIN-13A1.
 GN KIN13A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Jamin S., Quach H., Fellous M., Bourgeron T.;
 RT "Identification of the human KINESIN-13a gene, homologous to
 RT Drosophila kinesin-73 and candidate for schizophrenia."; submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AJ291578; CAC20442.1; -.
 DR HSSP; P33176; 1862.
 DR InterPro; IPR000253; FHA_domain.
 DR InterPro; IPR001752; kinesin.
 DR Pfam; PF00225; kinesin.1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00240; FHA.1.
 DR SMART; SM00129; KISC.1.
 DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2.1.
 DR ATP-binding; Coiled coil; Microtubules; Motor protein.
 KW SEQUENCE 1757 AA; 197084 MW; 92277256f8ddef0f CRC64;

Query Match 36.4%; Score 1465; DB 4; Length 1757;
 Best Local Similarity 43.6%; Pred. No. 2.7e-76;
 Matches 343; Conservative 119; Mismatches 205; Indels 120; Gaps 19;

QY 7 IKVYRVFPFARNREIDRGAKCIVRMENQTLTPPGAEEKARKSGKTIMGPKAFADR 66
 DB 6 KVAVRVAPNMRRELELTKCVMEGNGQYVLPSPSTKGEEK-----PPKVFADFY 59
 QY 67 SYWSEDK NARNYARQEDLPDGLVPLIDNAFKYGNCFAYGOTSGSKYSMMGYKEH 125
 DB 60 CFWSDESNITTKYAGQEVFCKLEGILEKAFQGNACIFAYGOTSGSKSFGMGHAQL 119
 QY 126 GVIRICODMFRINELQKDNLCCTVSVSLTYNEVRDLNP-STKGINKVAREHST 184
 DB 120 GLIPRLCALFKRIS-LEQNSQTFKVEVSWEIYNEVRDLNPCKGROSILKVAHEKVL 178
 QY 185 GPYEDLAKLVVRSFOETENLMDGKNKARTVAATNMNETSSRSHAVFTLTQKHDEET 244
 DB 179 GPYDGLSQLAVTSPEDIESLMSGNKSRTVAATNMNESSRSHAVFTLTQKHDEET 238
 QY 245 KMDTEKAKISLVNLAGSERATSTGATGARLKEGAELNRSLSLIGRVIAALADSSGKQ 304
 DB 239 GNSGEKSKVSLVNDLAGSERYSKTGAAGERLKEGSINIKSSTLTGLVYSSLDQAAGK-G 297
 QY 305 KNOLVPPYDSVLTWMLKSLGNSMTAMIAISPADINFEETLSLTVRADSKARKKNAY 364
 DB 298 KSKVPPYDSVLTWMLKSLGNSMTAMIAISPADINFEETLSLTVRADSKARKKNAY 357
 QY 365 VNEDPNAMIRELKEELAQLRSKLQSSGGGGGAGSGGPEESYPPDTPLEKQIVSIQ 424
 DB 358 VNEDPNAMIRELKEELAQLRSKLQSSGGGGGAGSGGPEESYPPDTPLEKQIVSIQ 384
 QY 425 PDATVKKMSKAEIYEQLNQSKLRLDNLQWEKLAETEEIHKEREALBELGISIEKGF 484
 DB 385 -----EAMKAELEKEEKESEKILKELTVTEEKLKTEETEAQERQLESMSIEMS- 438
 QY 485 VGPYHSKMPHLVNLSDDELLAEGLYVNIKRGQTRGVGNVNDQAEITLNSKILKEKCT 544
 DB 439 -GIYVGDDKCYLVNLNDPALNELLYTLK-DHTRVG---ADTSDDIQLFGIGQPOHCE 493
 QY 545 FEYV-DNVTITVNPKEKAVVNGVRIIDKPTRLRSGRYITLDFHIFRNHEEAREAEQ 603
 DB 494 IDIASDGOVTLTPKRNEMASCYNGTLVCGSTQLMGHDIRLIMGNNHFRILNKKRRRWMLK 553

QY 604 OSILRHSTNSQLGSPAPGRHRTLSKAGSDAGDSRSDSLPHPRKGSQMPYARREA 663
 DB 554 D-----FEKETGPP-----EHD-----LDAASEASSE-----PDYNEFEAOMEVI 588
 QY 664 SAILGLDOKISHLFDDELDAFPDQKARAVRGRGVEINE-----SDSOGSSFPV 713
 DB 569 MKTILNSNDPQNVN-QVLEKQYLEKRSALAEQRLMYRELEQLRQQLSPRQPOSSGPD 647
 QY 714 RDKYMSNGTIDNFSLDVIAITMGPTRSDDDDALEFGDKSKODASNVYVELRQQAQM 773
 DB 648 RLAYS-----QTA-----OQVYQMAEERD-ELFRQSLAKL 678
 QY 774 EALKTA 780
 DB 679 REQLVKA 685

RESULT 8

Q9H1H8 PRELIMINARY; PRT; 1770 AA.
 AC Q9H1H8;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE KINESIN-LIKE PROTEIN RBKIN2.
 GN RBKIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=RETINA.
 RA Chen D., Squire J.A., Gallie B.L.;
 RT "Genomic amplification in retinoblastoma narrowed to 1.2 Mb on
 RT chromosome 6p containing a novel kinesin-like gene, RBKIN."; submitted (Nov-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AY014404; AAC38891.1; -.
 DR HSSP; P33176; 1862.
 DR InterPro; IPR000253; FHA_domain.
 DR InterPro; IPR001752; kinesin.
 DR Pfam; PF00225; kinesin.1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00240; FHA.1.
 DR SMART; SM00129; KISC.1.
 DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2.1.
 DR ATP-binding; Coiled coil; Microtubules; Motor protein.
 KW SEQUENCE 1770 AA; 198550 MW; D926CDD8ED8CB31C CRC64;

Query Match 36.3%; Score 1463.5; DB 4; Length 1770;
 Best Local Similarity 45.5%; Pred. No. 3.3e-76;
 Matches 330; Conservative 116; Mismatches 189; Indels 91; Gaps 16;

QY 7 IKVYRVFPFARNREIDRGAKCIVRMENQTLTPPGAEEKARKSGKTIMGPKAFADR 66
 DB 6 KVAVRVAPNMRRELELTKCVMEGNGQYVLPSPSTKGEEK-----PPKVFADFY 59
 QY 67 SYWSEDK NARNYARQEDLPDGLVPLIDNAFKYGNCFAYGOTSGSKYSMMGYKEH 125
 DB 60 CFWSDESNITTKYAGQEVFCKLEGILEKAFQGNACIFAYGOTSGSKSFGMGHAQL 119
 QY 126 GVIRICODMFRINELQKDNLCCTVSVSLTYNEVRDLNP-STKGINKVAREHST 184
 DB 120 GLIPRLCALFKRIS-LEQNSQTFKVEVSWEIYNEVRDLNPCKGROSILKVAHEKVL 178
 QY 185 GPYEDLAKLVVRSFOETENLMDGKNKARTVAATNMNETSSRSHAVFTLTQKHDEET 244
 DB 179 GPYDGLSQLAVTSPEDIESLMSGNKSRTVAATNMNESSRSHAVFTLTQKHDEET 238
 QY 245 KMDTEKAKISLVNLAGSERATSTGATGARLKEGAELNRSLSLIGRVIAALADSSGKQ 304

Db 239 GNSGKSVKSVSLVDLAGSERVSKTGACERLKEGNSINKSLTTGLVLSLADQAAGK-G 297
Qy 305 KNOIYPYRDSVLTWLLKDSLKSGNSMTAMIAISPADINFEETLSTLRVADSARKIKNHAV 364
Db 298 KSKFPYRDSVLTWLLKDNLGNSTSMIATISPADNYEETLSTLRADRAKRIYNHAY 357
Qy 355 VNEDPNAMIRELKEBELAQLRSKLSQSGGGGAGCGGPGVEESYPPTPLEKQIVSIQO 424
Db 358 VNEDPNAMIRELKEBELAQLRSKLSQSGGGGAGCGGPGVEESYPPTPLEKQIVSIQO 424
Qy 425 PDATYKKMSKAIYVQLNOLSEKLYRDNLQWTEERLAKTEEIHKEREALLEEKISTEKG 484
Db 385 -----EAKAPLKEKELESEKLELKTETWEKLRKTEEIAQEROKOLSESGISLEMS- 438
Qy 485 VGPYHSEKMPHLVNLSDPLAECLVYNIKPQTRVGNVNDTOAIEIRLNGSKILEKHC 544
Db 439 -GIRYGDCKCYLVNLMDPALNELLYLK-DHTRVG---ADTSODIDQFGIGIQPQICE 493
Qy 545 FENV-DNYVTIVPNEKAAMVNGVRIDKPTRLSGYRIILDFHIFRNHPPEARAEOR 603
Db 494 IDIASDGVTLTPKEMARSCVNGTLCSTQIMHGDRIILMGNNHFRINLPKRKRDMWK 553
Qy 604 OSLRHSTVNSQLSPAPGRHRTLSKAGSDADGSRSDPLPHRGKDSMDFVARRAEA 663
Db 554 D-----FEKETGPP---EHD-----LDASEASSE-----PDYNTFEAOMEVI 588
Qy 664 SAILGLDQKISHLTDELDAFLPDVQKARAVRGGLVEDNE-----DSDSQSSFPV 713
Db 589 MKTLSNPNVQVNV-QVLEKQYLEEKRSALAEQRLMYERELEQLRQQLSPDQPOSSGPD 647
Qy 714 RDKYMS 719
Db 648 RLAYSS 653

RESULT 9
Q9NH9 PRELIMINARY; PRT; 1805 AA.
AC Q9NH9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE KINESIN-LIKE PROTEIN RBKIN1.
GN RBKIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RA Chen D., Squire J.A., Gallie B.L.;
RT "Genomic amplification in retinoblastoma narrowed to 1.2 Mb on
RT chromosome 6p containing a novel kinesin-like gene, RBKIN.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AI014403; AAC38890.1; -
DR HSSP; P33176; 1B62.
DR InterPro; IPR000253; FHA_domain.
DR InterPro; IPR001752; kinesin.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PRO0380; KINESINHEAVY.
DR SMART; SM00240; FHA; 1.
DR SMART; SM00129; Kisc; 1.
DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW ATP-binding; Coiled coil; Microtubules; Motor protein.
KW ATP-binding; Coiled coil; Microtubules; Motor protein.
SEQUENCE 1805 AA; 202258 MW; AAF249B859A04B7A CRC64;

Query Match 36.3%; Score 1463.5; DB 4; Length 1805;
Best Local Similarity 45.5%; Pred. No. 3.4e-76;
Matches 330; Conservative 116; Mismatches 189; Indels 91; Gaps 16;

Qy 7 IKVAVRVPFNAIREIDRAKCIIVRMEGNOTILTPPGAEKARKSGKTIIMGPKAFADR 66

Db 6 VKAVAVRVPFNAIREIDRAKCIIVRMEGNOTILTPPGAEKARKSGKTIIMGPKAFADR 59
Qy 67 SYWSDK-NAPYARQEDLFODLGPILDNARKGNINCFATGQSGSGVSMQYKREN 125
Db 60 CFWSDKSTTKYAGQEVVFKCLGRIEKAFOGNACIPAFQGTGSGSFSMGMHARDL 119
Qy 126 GVIPRICODMFRINELQDKMIMCTVEYSYLEIYNERVRLAN-STFGKGLVKEHPST 184
Db 120 GLIPRLCALFKRIS-LEQNESQTFKVEYSYEITNEYRDLDDKSGNSQSLKVEHVL 178
Qy 135 GPYEDLAKLVVRSPOEILNIDEGNKARTVAATNNETSSSHAVFTLTLOKHDEET 244
Db 179 GPYVGLSOLAVTSFEDIESLMESEGNKSTVAATNMNESSRSHAVFTLTLOTDLOS 238
Qy 245 KMDTEKVKISIVDLAGSERATSGATGARLKEGAENIRNSTGLGRVIAALADMSGKQ 304
Db 239 GNSGKSVKSVSLVDLAGSERVSKTGACERLKEGNSINKSLTTGLVLSLADQAAGK-G 297
Qy 305 KNOIYPYRDSVLTWLLKDSLKSGNSMTAMIAISPADINFEETLSTLRVADSARKIKNHAV 364
Db 298 KSKFPYRDSVLTWLLKDNLGNSTSMIATISPADNYEETLSTLRADRAKRIYNHAY 357
Qy 355 VNEDPNAMIRELKEBELAQLRSKLSQSGGGGAGCGGPGVEESYPPTPLEKQIVSIQO 424
Db 358 VNEDPNAMIRELKEBELAQLRSKLSQSGGGGAGCGGPGVEESYPPTPLEKQIVSIQO 424
Qy 425 PDATYKKMSKAIYVQLNOLSEKLYRDNLQWTEERLAKTEEIHKEREALLEEKISTEKG 484
Db 385 -----EAKAPLKEKELESEKLELKTETWEKLRKTEEIAQEROKOLSESGISLEMS- 438
Qy 485 VGPYHSEKMPHLVNLSDPLAECLVYNIKPQTRVGNVNDTOAIEIRLNGSKILEKHC 544
Db 439 -GIRYGDCKCYLVNLMDPALNELLYLK-DHTRVG---ADTSODIDQFGIGIQPQICE 493
Qy 545 FENV-DNYVTIVPNEKAAMVNGVRIDKPTRLSGYRIILDFHIFRNHPPEARAEOR 603
Db 494 IDIASDGVTLTPKEMARSCVNGTLCSTQIMHGDRIILMGNNHFRINLPKRKRDMWK 553
Qy 604 OSLRHSTVNSQLSPAPGRHRTLSKAGSDADGSRSDPLPHRGKDSMDFVARRAEA 663
Db 554 D-----FEKETGPP---EHD-----LDASEASSE-----PDYNTFEAOMEVI 588
Qy 664 SAILGLDQKISHLTDELDAFLPDVQKARAVRGGLVEDNE-----DSDSQSSFPV 713
Db 589 MKTLSNPNVQVNV-QVLEKQYLEEKRSALAEQRLMYERELEQLRQQLSPDQPOSSGPD 647
Qy 714 RDKYMS 719
Db 648 RLAYSS 653

RESULT 10
Q9NOT8 PRELIMINARY; PRT; 1826 AA.
AC Q9NOT8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE KINESIN-LIKE PROTEIN GAKIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-20435838; PubMed-10859302;
RA Hanada T., Lin L., Tibaldi E.V., Reinherz E.L., Chishti A.H.;
RT "GAKIN: a novel kinesin-like protein associates with the human
RT homologue of the Drosophila discs large tumor suppressor in T
RT lymphocytes";
RL J. Biol. Chem. 275:28774-28784(2000).
EMBL; AF279865; AAF81263.1; -.

Thu Jul 25 08:37:55 2002

us-09-235-416-1.rsp

Page 12

```

QY      643  -----SPLPHRGKDSMEFYARREASATLGLDOKIS-HLTDELAHLFDDYQKARAVR 696
Db      544  EEELMSTLP-----LVORANMATELGRNVKFEIVLVSPEM-----RGLTIS 654
QY      697  GLVE 700
Db      635  GLTE 638

```

Search completed: July 25, 2002, 05:48:42
Job time: 256 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 03:47:27 ; Search time 3312.9 Seconds

(without alignments)
14856.837 Million cell updates/sec

Title: US-09-235-416-2

Perfect score: 2352
1 atgtcggcggtggaatat.....aaacagcgaagcaggaattc 2352

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_dat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	474.8	20.2	6972	9	HAAXTRSYV	X90840 H. sapiens m
2	463.8	19.7	7401	3	AF247761	AF247761 Drosophila
3	437	18.6	5421	10	MUSKIF	D29951 Mouse mRNA
4	421	17.9	5739	10	AB023656	AB023656 Mus muscu
5	419.4	17.8	4862	10	MUSKIFB	D17577 Mouse mRNA
6	398.4	16.9	5575	9	AY043362	AY043362 Homo sapi
7	398.4	16.9	6816	9	AF257176	AF257176 Homo sapi
8	382.2	16.2	4082	9	HS091329	U91329 Human kines
9	382.2	16.2	6217	9	AB014606	AB014606 Homo sapi
10	361.8	15.4	4117	9	RNAJ696	AJ000696 Rattus no
11	361.6	15.4	3303	10	AB074017	AB074017 Mus muscu
12	352	15.0	5170	3	CELU104A	M58582 Caenorhabdi
13	352	15.0	7107	3	DM081788	U81788 Drosophila
14	346.6	14.7	6618	10	AB037923	AB037923 Mus muscu
15	328.2	14.0	5700	9	HSR291578	AJ291578 Homo sapi
16	328.2	14.0	5773	9	AY014404	AJ291578 Homo sapi
17	328.2	14.0	5780	9	HSR291579	AJ291579 Homo sapi
18	328.2	14.0	5878	9	AY014403	AY014403 Homo sapi
19	318.8	13.6	9924	10	AF090190	AF090190 Mus muscu
20	317.2	13.5	5743	10	AF131865	AF131865 Mus muscu
21	316.6	13.5	8776	9	AB017133	AB017133 Homo sapi
22	316.2	13.4	10682	6	AX039604	AX039604 Sequence
23	314.6	13.4	8743	9	AF279865	AF279865 Homo sapi
24	311	13.2	6325	3	AF245277	AF245277 Dictyoste
25	307	13.1	561	8	CNS01C94	AL115232 Botrytis
26	300.6	12.8	2067	10	AF083331	AF083331 Rattus no
27	241.8	10.3	5532	3	AY070884	AY070884 Drosophila
28	235.4	10.0	2819	3	AB023497	AB023497 Caenorhab
29	215.6	9.2	130049	8	NC17E5	AL513467 Neurospor
30	193.4	8.2	3432	10	AB008667	AB008667 Mus muscu
31	192.6	8.2	2282	10	MOSKIF3	D12645 Mouse mRNA
32	192.4	8.2	2213	3	SUSPKIN2A	L16593 Strongyloce
33	189	8.0	6586	6	AX333633	AX333633 Sequence
34	189	8.0	6586	6	AX334308	AX334308 Sequence
35	189	8.0	6586	6	HUMORFW	D26361 Human mRNA
36	189	8.0	172148	3	LMRP696	AL512293 Leishmani
37	188	8.0	451	9	AF004425	AF004425 Homo sapi
38	187.2	8.0	4021	9	AB046810	AB046810 Homo sapi
39	184.2	7.8	2391	3	AF149285	AF149285 Caenorhab
40	178.2	7.6	69523	2	AC094844	AC094844 Rattus no
41	177.8	7.6	3319	3	LEIKINLIKE	L07879 Leishmania
42	177.8	7.6	3319	6	AR071970	AR071970 Sequence
43	177.8	7.6	3319	6	I11729	I11729 Sequence 2
44	177.8	7.6	3319	6	I88957	I88957 Sequence 1
45	176.8	7.5	453	5	MSU34655	U34655 Morone saxa

ALIGNMENTS

RESULT	1	HAAXTRSYV	6972 bp	mRNA	linear	PRI 28-MAY-1996
LOCUS	HAAXTRSYV	H. sapiens mRNA for axonal transporter of synaptic vesicles.				
DEFINITION	X90840	X90840				
ACCESSION	X90840.1	GI:1212916				
VERSION	ATSV gene; axonal transporter of vesicles.					
KEYWORDS	human					
SOURCE	human					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.					
TITLE	Furlong, R.A., Zhou, C.Y., Ferguson-Smith, M.A. and Affara, N.A.					
JOURNAL	Characterization of a kinesin-related gene ATSV, within the					
MEDLINE	tuberosus sclerosis locus (TSC1) candidate region on chromosome 9034					
REFERENCE	Genomics 33 (3), 421-429 (1996)					
AUTHORS	96299637					
TITLE	2 (bases 1 to 6972)					
JOURNAL	Furlong, R.A.					
COMMENT	Submitted (16-AUG-1995) R.A. Furlong, University of Cambridge, Dept					
	of Pathology, Tennis Court Road, Cambridge CB2 1QP, UK					
	Overlaps with M78444, M78705, T07754, T15633 and T77291.					

1578 tacacaacggaacatcgtcgtacgaagtcctcgaagaacacgtgacgttga 1637
1635 GAGGCGGACGACATGTTCTGAGTGGGACATTCACAGAGGAGGACATGCTTCG 1694
1638 aaatgtgacaac-----gtgtgacacatcgtgcgaacgaagagctgc 1682
1695 GAGCAGCTCCAGGAGGACGACGAACTGCTGCTGAGCCCTTGAGGGGACACA 1754
1683 tgcattgtgaacgcgtgcgaatcgacaaagctactcgcctccgcaaggttaagat 1742
1755 CACCTACGTCATGTCAGCAAGAAAGTCACAGAGCCACGATCTGCTCAGCAACCGCAT 1814
1743 catcttggtgacatccacatcttcgaatccacacacgcaagaagacgtgcgaag 1802
1815 CATCATGGGTAAAGACCATGCTGCTCCGGTTCACCAACCCGAGAGGCCGAGAGG 1874

RESULT 2
AF247761 7401 bp mRNA linear INV 05-JUN-2000
LOCUS Drosophila melanogaster kinesin superfamily member Dunc104 mRNA,
DEFINITION complete cds.
ACCESSION AF247761
VERSION AF247761.1 GI:8248420
KEYWORDS fruit fly,
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 7401)
REFERENCE Grosberg, R., Saxton, W.M. and Dickson, B.J.
TITLE Characterization of the Drosophila Unc104/TrfA homolog, Dunc104
JOURNALS Unpublished
2 (bases 1 to 7401)
AUTHORS Grosberg, R., Saxton, W.M. and Dickson, B.J.
TITLES Direct Submission
JOURNALS Submitted (23-MAR-2000) Institute of Molecular Pathology, Dr.
Bohr-Gasse 7, Vienna A-1030, Austria

FEATURES
source location/Qualifiers
1..7401
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2"
/map="53D"
431..5446
/product="Kinesin superfamily member Dunc104"
/protein_id="AA074192.1"
/db_xref="GI:8248421"
/translation="MSSVKVAIVRVRPENSREIARESKCIIMAGATTAITNPKVPNT
SDSVKRFNFDYSYSHDHADPSTOSVYVDIGEEMLGSPGYNVICERVDILNP
KSYTMGROEQOQGIIPMTCKDLPTRIQDETDLKYSVSVMEITCERVDILNP
KNGKMLVREHPLGLPTVEDLSKLAIVDYQDIDHIDGNAKTVAAANMETSRSRSH
AVPTLFTQRHDLMTNLTETKSKISLVLDAGSERDSTAKTRKLEGNINIKSLT
TLGKVALAEVASAKKRNKADFIPEPDSALMLRENGKSLDKMAIAISPAID
NYDETLSTLRADRAKQIVCAVVEDANALIRELKEIDILKAEIGEOED
ELTKESTVKSPTKSRNGSTTEMAVDOLASEKLIASELTWEKIKRETEINORE
AVPEMGVAKEOSITVGVSPKTPHVLNENPNISECLYITIKBSLTGLRTHAN
VPDIOISGHILHECTFENKSTVTLLPKKDAIIVNGKLVPEVLKGLSRVILG
KNHFRFTNPQARELBDKIETENEAEVEKRTDQVDMNPAOCELLEKQIDILAK
MKRLDLLEQYKREKLQAOQEEQKRTYARIDALQYVEQSMYSKSSPED
PHQEDVYTNPMYSCWTARAGLAANAFKRWYRHOPLSLRDLDMGNAIFLKEMATIS
VELKKVQFOFTLITDLSPLPELAISTVA PVHQBDEFGAPRPSKTVIAVEYDTK
GATHMSLEKIROLELMREMYHNEAEKSPSPRYNESLIGGDFPDIFPFRVGR
SEITLSNLTLPVLNKAIVAVNERGDVGYLRLAVDVLDEESIDFNNGVQSAALV
NEDAKPKRYALNKEDVORYTNGDLSKLELEDVDSGIDNSASBECHENSEEP
GEHLQVKEEFRTVTLQATGIAEYADIFCOFNFLHHEEAFSTEPYKNSGAPLG
FYHQNTLVPTKSFIEYKLPQIMFKIFGHOYHPLHKAKODFVSPPRMPLPS
IPISQVPSPEKGPLPCAPSTVLAKHDVILWFEICELAPNGEYVPSVEISDLPCR
GLPILHGIQRIKRIITTVHEPTTEYKMKDINELVAGRIANTPESSDEDDACVLSIG
LPEGALEVPQDNRSPYRFEAMDSILHNSALLNRVSGSGFTIITISAYLENCAR

PAITTKDLSMVIYGRDARTGPRSLKHLFSGOYRNPBEANRLTGVETALIRASEAGSPG
VORROARVLDSTSTYVRGENWLEHGRPRDSILIPHOWLEKTLREVEGRHRLILL
BERLMDPNPNTETKREKYNCLARARPSYKPHMYIPQSPVPVDPQOITEREYNO
ROODMLKCIKIVQGRYTRISEANDOTOSDVSDBGADMTVSCISSMNNKFEVI
RRRLCSPPDADAPNWEAPAPITOPALPLRLVPELEIRVSPVARKGLVLEHGG
SOWKRWYIVRRPVYFIVRSEDDPVERAVLNTATPVAHEVCSQOAMVVKIPNFSVYTK
HRGYLQTLGDKREVHDWLYAIINPLLAGQIKSRLRRTLEPASQTSIQATNAANNS
ASK"

BASE COUNT 2120 a 1714 c 1801 g 1766 t
ORIGIN

Query Match 19.7%; Score 463.8; DB 3; Length 7401;
Best Local Similarity 57.2%; Pred. No. 8.7e-121;
Matches 1033; Conservative 0; Mismatches 687; Indels 87; Gaps 7;

20 tcaagtgatgtgtgtggtgacgcgcgttcaacgcccagagaatcgacgtgagcaaat 79
441 TTAAAGCTGGCGGTGCGCAAGTGGCCGCCCTTCACATCGCGCAATAATACAGGAGTCAAAAT 500
80 gtaattgtgcgagtggaaggaatcaacacatccctcaacccctccctccggtggtccgaagaga 139
501 GCATCATCGAGATGGCAGAGGCCACACAGCCATCACCAATCCAAAGTGCCGCCACACA 560
140 aggcgcgtaaagtgycacaaactatattgattgtgcccgaagagcatctgcgtcgtatcgt 199
561 CA-----AGCGATTCACTAAAGCGCTTCACTTGATTACT 596
200 cgtattgtccctttgacaaagaatgctcccaactatgcgagacgagaagacattatccaaag 259
597 CCTACTGGTCACATGATGATCACACAGATGCGGATTTCTCCACACAAATGATGATGATACAG 656
260 atctcggaatcccgcttcgtgataagatgaaggttaaggttaacattgattatcttgcct 319
657 ACATTGGCAGAGATGATTACAGCATCTCTCGATGGCTACCAATGTCGTATCTTGTGCT 716
320 acggtcagaacgcgtctcggaagtcctatccaatgataggcgtatggaaggaagc----- 373
717 ATGGCCAGACTGGAGAGTGGAAAGTCTTATACATGATGGCCAGGACGAGGACACACAG 776
374 atggcgtatcccgcggaatttgcgaagacatgctccgagcgtatgaatgaatgcgaagga 433
777 AGGGCATTAATACCAAGATGATGACAGATGCTGTTCACTGCGATACAGGATACGAGA--- 833
434 acaagaacccctcactgtgcacgcgtcgaaggttcgtcacttggaatattcaatgaacgcgtgc 493
834 CCGATACCTCAAGTACTGCTGTGAGGTTCTTATATGGAATCTATTTGCGACCGAGTCA 893
494 ggaactgtgtaatccgtcgcgaacaaagggaatccaaggtcccgagacacccgctgcagcg 553
894 GGGATTGTGTAATCCGAATAACAGGAGCAATGCGGTGAGGAGCATCCCTTTTGG 953
554 gccctcagtggaagacttgcggaagcgtgctgcgtacatccaagaatcgaaatc 613
954 GTCCTATGTGAGGATTTGTCACAACTAGCCCTTACCATTCACGAGCATTCACAGATC 1013
614 tcatgataagggaacaaagccagaagcgttgcgcgcgaacaaagaaagaaatcca 673
1014 TCATTGATGAAGCAACAAAGCTCGAAGCTGTGCACTCTTCAACAGAACGAAACCAAGTT 1073
674 gtcgaatccacgcgcgtctcactttgaccttgacgcgaacaaagtgcatagtgaagagacca 733
1074 CCGGCTCTATGCGCGCTTTTACAAATTTTTCACCAAGTGTCTCATGATTCGATACCA 1133
734 aaatgacacagaagaaggttcggaagatacagtcctgtagaattgcggggttcgtgaagcgag 793
1134 ATTTCACCAACAAAGATATCCAGATTAAGCTTGATAGCTGCGCGGTGCGAAGCAG 1193
794 caagtcacacggagactactgtagcgcgactgaaggaaggtgcgaagataccaacgcgtcac 853
1194 CGGATTTCACGCTGTGCAAGGGGCACTGCTTGAAGGAGGAGGACCAATTAACAAAGTCTT 1253
854 ttctgacacctagtgctgtgattgcagcgctagcgatatagtctgcgtggaacagaaga 913

Db	1254	TGACCA	CACTTGGGCAAGATTATCTACAGCTTGGCGGAAGTGTGCTGTTCCAAAGAA	1313
QY	914	agaatcag	-----ttgtaacctaccgagatccgtaactgacgtgtcctcga	964
Db	1314	ACACCA	AAAGGAGATTTATTCCTCGTACCGTATTCGGCTTGAACCTGGCTGTACGTG	1373
QY	965	actccttggagagcaactcgatgacccgcaatgtgcgaatttcgcctgtctgata	1024	
Db	1374	AAACTTGGGAGGAATTCAGACGACGTATGATTTCCGGCTATCTCACCAGCAGTATTA	1433	
QY	1025	acttgaagaagactctcaagtaaccttcogataatgacgactctgcgaagcgaa	1084	
Db	1434	ACTAGAGTAAACCTCAGACACACTGGCTATCCGATGTGCCAAAGAAATGTGTGA	1493	
QY	1085	acgcagctgtgtcaatgaagagaccggaacgcgaggtgtctcggaattgaagag	1144	
Db	1494	AGGCTGTGTGTAATGAAGATGCCAAATGCCAAGCTGATTGCCAACTCAAGGAGGATCC	1533	
QY	1145	cgcagctgaagagcaaaactccagacagctgtgtgaaagtgtgtcgaagagctgtc	1204	
Db	1554	AGAAAGCTGCGGAGCTTACTCA	-----AGCGAGGGCCANTG	1589
QY	1205	gcgagccaggtgaggaatcgtaccgcgcgacacgcgcctcgaagaacgaatcgtgtcga	1264	
Db	1590	AAGTGCAGGAAGAGAGTACCTCACAAAGTCAACGGTGTATCAAGTGGCCACTAAGATAC	1649	
QY	1265	ctcagcagccgagatgcgcacgtaagaanaatgacgaagcagaatctgtgagcaactga	1324	
Db	1650	GTATATC-----GTATGTGATCCACCCAGAGATGGCTGTGGATCAAGCTGC	1694	
QY	1335	accagagtgagaagctctatcggatctcaatcagaacctgggaagagaagctgtgcgaaga	1384	
Db	1695	AGGCGAGGAGAACTTATTTGACAGACTCAACAGACCTGGGAGGAGAACTTAAGCGCA	1754	
QY	1385	ccgaggaatccacaagaagacggaagcgcgctcgaagaagctgtgtatcagcatcgaa	1444	
Db	1755	CCGAGGAATTCGCTGCTACGCGAGAGCGGTCTTCGCCGAATATGGTGTGGCTGTCAAGG	1814	
QY	1445	agg-----gcttgttcgtgcctctaccactccaaagaatgacacatctagtcgaactga	1498	
Db	1815	AAGATGGAATTAACAGTTGTGGCGTATTTCTCAACCGAAGAAAGACTCCGATTTGGTCAACCTAA	1874	
QY	1499	gcgagtatcctctctctgtgcgtgaatgtctctgtcttaacaactaaagcccgaggcagaagaag	1558	
Db	1875	ACGAGGATCCCAATCTTGTGAGTGTGCTTACTCAATCAAGGAAGGCTTAACTTCGCT	1934	
QY	1559	cttgaagaactlcaacaagaagatacacaagcggaaatcgtctgaagcgttccgaagactcgga	1618	
Db	1935	TGGGTACCCATGAAGAAATGTGCCACGAGACATTTACGATCTCCGGATGTGCACATCTCA	1994	
QY	1619	aagaacactgtacgttctggaanaatgtggaacaagctgtgtgaacatcgtgtccaaacgaagaag	1678	
Db	1995	AAGGACACTGTGACCTTTGGAAMCAAAAGCAAGCGGTTTACCTGTGTCACACAAGGATG	2054	
QY	1679	ctgcgttcatactgtgaacgggtgtgcgatacgaaagctactcgtcctcgcagcggtcata	1738	
Db	2055	CTATCATCTTATGTAAATGAGACGCAAGTGTGTGAACCGGAGGTTCTTAAACACGGTTTC	2114	
QY	1739	ggatcactcttgggcgatttccacatcttttcgattcacaactccggaagaagactcgtgcg	1798	
Db	2115	GGGTGATCTCTCGGAAAGAACCAAGGTGTCGGCTTTACCAATCAAGACAGGAGCGGAT	2174	
QY	1799	aacgcga	1805	
Db	2175	TACGGGA	2181	

RESULT	3	
MUSKIF		
LOCUS	MUSKIF	5421 bp mRNA linear ROD 07-FEB-1999
DEFINITION	Mouse mRNA for kinesin family protein kif1a, complete cds.	
ACCESSION	D29951	

VERSION	D29951.1	GI:976234
KEYWORDS	kinesin family protein KIF1a; monomeric motor; neuron-specific kinesin superfamily motor; axonal transport; organelle transport. Mus musculus Brain cDNA to mRNA.	
SOURCE	Mus musculus	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 5421)	
AUTHORS	Yamazaki, H.	
TITLE	Direct Submission	
JOURNAL	Submitted (15-APR-1994) Hiroto Yamazaki, University of Tokyo, Faculty of Medicine, Department of Anatomy and Cell Biology, 7-3-1, Honjo, Bunkyo-ku, Tokyo 113, Japan (E-mail:Yamazaki@kinesin.kaiho1.m.u-tokyo.ac.jp, Tel:3-3812-2111(Ex.3336), Fax:3-5689-4856)	
REFERENCE	2 (sites)	
AUTHORS	Okada, Y., Yamazaki, H., Sekine-Aizawa, Y. and Hirokawa, N.	
TITLE	The neuron-specific kinesin superfamily protein KIF1a is a unique monomeric motor for anterograde axonal transport of synaptic vesicle precursors	
JOURNAL	Cell 81 (5), 769-780 (1995)	
FEATURES	95292344	
FEATURES	Location/Qualifiers	
SOURCE	1. .5421	
FEATURES	/organism="Mus musculus"	
FEATURES	/db_xref="taxon:10090"	
FEATURES	/tissue_type="Brain"	
FEATURES	121. .5208	
FEATURES	/note="121-2122: highly homologous to KIF1b; 331-1174: kinesin type motor domain; 4834-5163: PH domain"	
FEATURES	/codon_start=1	
FEATURES	/product="kinesin family protein KIF1a"	
FEATURES	/protein_id="BAA06221.1"	
FEATURES	/db_xref="GI:976235"	
FEATURES	/translation="MAGSASVAVYVRPPRREMSRDSKCIQMSGSTTITVNRQKRETPSPFSFDISTWSTSPEDINIVASQKQVNRDIEEMLQHAFFESYNYCITRAYGGTGAGKSYTMGQKEKDQGGIIPOLCEDSLFASINDTNMGSYSEVSFMEIYCEVRDLNMPKKNKRLNRREHLPDIPYVEDLSKLAIVSYNDIIDLMSGKPRVATNNMTSSRSHAVNFIIFLQKHDAETNTITTEKVASIIVLDAGSEKSDGKATKEGANKIKLSLTLGIVSIALMADESGPNKKKKTKTDLPYVSDSLVWMLIRPNIIGNSRTVMVAISPRADINDVETLIRVADRRAKQITPCNATINDEPNKTIIRLQKQVYPRDLIYLAQGLDITDMTALVGMSSSSLALSRSAAVSSTHEILFAPSSSEATIEYKLEKTIYALNETWEKRLRTETAKREREALMAGVAMRDEGDTLVSFSPKTRPLVNLNETDMSCELLYVNGKIVTPSILRSNGRIIMKSHVFEFNHEEQARORERPCAFETAPEDVAFQAFORBLKQOIKEMQKEORLOLEDOYRREBEATATLQORJLDYRSKLEALQKQKDSRYTYPEDEEEDPEEDVOWTHERCELATLNAFKWQYOTSLRDLMLGALFLEKNAISVELKKYQFOFOLLTDLTSLPLPDLIPREAKQRETPRRTIVAYENDOKGATHYMTLEKLRQRLDIRLREMTDRAAEYPPSSVEEDCWVYTGPPFDTRPFWFRFLVAGTFLVLSNLIVPPLVRAIVASEKEVGKFLVAQVQIASDEADPDSGVRSQSTAKTFFVDQHEKRFQSESCPVAGMSRSTQSEELIRVGOQADAGASADENVNNKCSAVPBGGSALMDSPEKRALDGPDLDTALDRLGISTFTFRVTVQAQSIISAEVADICQFNEIHRHDEAFSTPELKRNGRPPLGFYVONALNAVETKSFIEYIKSPQIVFEVFGHOHPRLPCKKDVSPLPSSRRHPFVAPWPLSKVPATKLSMTPRSGPQCKYDLVYEICELANNGDYIIPAYDHRGAMGTGFIHOGIORTIYTLITLHETSGSHIRKWEKREIVLYGRIRNPEETDEALIDPNTLSLNIITSSGYNHPAQRDVRFPFGNDTRTFYFPAEMDSMNSILNAYTPIREKTYMTLSAIIEMENCQIAPVITIKDCQWFSYSDAKIPARSILRNLFSGSLIAYTEGNRVGYVELSLCHVADAGSPGMQRRRRVVDITSAAYNGEENLAGMRRSQSLITLTDQHELEKLSLTLSEVETKRTHILREKLETFROGPPVLSPASEDESSSSGASAPLSAEGSPLEDAKPNRORRELAVKCLRLIMHTFENEYTHSHVICASDESKEMVTLMLDSGMSPLGATITPPSSCPCLIGRGYAGDVTRPOCSPRASEPELLDELQSKTMSPVATKTEKREKQRLVLPDIOLEIVSPSYSKGTLHLLEPTATMAKRFVYVRPYYAYMNSDKITVEVFVNLNTSQVETSEDQAMAKITPNTFAVCTEHRGILLIANSMDKMDHWLVAFNPLLAGTITRSKLSRRAESDMRY"	
BASE COUNT	1324 a	1593 c 1467 g 1037 t
ORIGIN		

Query Match	18.6%	Score 437;	DB 10;	Length 5421;
Best Local Similarity	56.2%;	Pred. No. 3.8e-113;		
Matches 1036;	Conservative	0;	Mismatches 725;	Indels 84;
				Gaps 8;

Db 123 GCGCCGGGCGCTCTGTAAGGTGGCGGTGCTCCGCCCTTCACTCCCGGGAATGAG 182
 Oy 66 ccgtgagcaaatgtatgtgagatgaaagaaatcaaatcaatccatccctcc 125
 Db 183 CGAGACATCCAAAGTGCATTCATGCTGTGAAAGCACACACTATTGTCAACCC--- 239
 Oy 126 ggtgtcgaaagaaagcgcgtaaaagtlygcaaaactaltatgalygcccgaagcalt 185
 Db 240 -----CAACAGCCCAAGAGACACCAAGAGCTT 269
 Oy 186 tgcgttcgactcgttcgtatgttgccttggac---aagaatgtcccaactaagcaagaa 242
 Db 270 CAGCTTTATTACTCTCTACTGTGCATACATCTCGCCGGGAGCATCAACTAATGACCTCTCA 329
 Oy 243 gaaagacctatccaaagctcggagtcggcttcgtgataatgacttcaatgaagtataa 302
 Db 330 CAGAGAGGCTTACAGAGGACATTGGGGAAGAGATGCTACAGCATGCCCTTGAAGGCTACAA 389
 Oy 303 caattgtatcttcgctacggttcagaccggttcgagcaagttcattcaatgatygtcta 362
 Db 390 TGTGTGATCTTGTGCTATGGGAGACGGGGCGGCAAGTCTCAACCATGATGGGGA 449
 Oy 363 tggcaag-----gagcatggtgtatcccggtatgtgcaagacatgttccggtat 416
 Db 450 ACAGAGAGAAGACACAGAGGATTCACAGCTTCGGAAGACCTCTCTCGAAT 509
 Oy 417 taatgaactgcaagaaagaaactcaactgtcacccgtcgaagtgttctgacttggaaat 476
 Db 510 CACAGCACACCCCAACGAC---ACATGTCTCTACTCTGTGGAGGTGAGCTACATAGGAGAT 566
 Oy 477 ttacaatgaagagtgagagactgtgtaatcgttcagacaagaagggaatcacaagttcag 536
 Db 567 TTACTGTGAGGAGATCCGTAACCTCTGTGAACCCCAAGAAAGGAACCTCGTGTGAG 626
 Oy 537 agaacaaccgttcagaccggtccctacgtgtgagagacttgcgaaggtgtgtgtcgatcalt 596
 Db 627 GGAGACACCTCTGCTGGGGCTTTATGTGAGAGACCTCTCCAAAGCTGGCTGAACCTCTCA 686
 Oy 597 ccaagaatctgaaatctcaatgagagggcaaaagcaaaagcggttgcgcccacaa 656
 Db 667 CAATACATCCAGAGACCTTCATGACTCAAGGAAACCAAGCCAGACTGTGGCCCAACAA 746
 Oy 657 catgaagagacatccagtcgaatcccaagcggttctcaacttgaacttgaacgcaaaagt 716
 Db 747 CATGATGATGACAGCAGACGCGTCTTCACGACGCTTCAACATCATCTTAAACCAAGAGG 806
 Oy 717 gcatgataagagagacaaatgtgacacagagaaggttgcgaagatcagtcgttagattt 776
 Db 807 CCATGATGACAGACCAACATCACACAGAGAAGTGAACCAAAATCAGCCTGGTGAACCT 866
 Oy 777 ggcgggttcgtgagcgcaagtcgacccgagctactgagcgacagtcgaagaggtgag 836
 Db 867 GGCAGGCGATGACAGACGCACTCCACAGAGACCCAAAGGCGCACTCAAGAGAGGGGCG 926
 Oy 837 agagatcaacccgtcacttgcacccctaagtcgtgtgattgtcagcgctgagcgatgtc 896
 Db 927 CAACATCAAAAGATCTTTGACCAACCTGGGAAGGTCAATCTGCTGGCTGGTGAATGA 986
 Oy 897 gtccgga-----aaacagaagaagaatcagttagtaacttaccagagattc 941
 Db 987 CTTCTGACCCCAACAAGAACAGAAAAAGAGACAGATTTTCATCCCTTACCGAGACATC 1046
 Oy 942 ggtactgagcggttcgtgaagagactccttggagggcaactcgaatgacgcatgtatgc 1001
 Db 1047 CCTATTGACATGCTTTCCTCCGGAAAAACCTCGGTGCACTCAAGAGACAGCCATGGTGGC 1106
 Oy 1002 cgcacatcgcgtgtgataacttgaagaagactcagtaaccttcgatalatgcgga 1061
 Db 1107 ACCCTTGAGACCTCGGAGACATCACTATGATGAGACCTTCACACACTAGATATTCAGAG 1166
 Oy 1062 cctctgaaagcaatcaagaacacggtgtgtcaatgaagcccgaaagcgcgatgtat 1121

Db 1167 TCGGCCAAACAAATCCGCTGCATGCCATTCATCAATGAAGACCCCAACAGCTGAT 1226
 Oy 1122 ccgcaggttgaaggaagaactcgcgcagctgaggaagcaaatcccaagcagtggtgag 1181
 Db 1227 TAGAGAACTGAAGGATGAGGTGACAGGCTGAGAGACCTGCTGTATGCCCAAGCTTGG 1286
 Oy 1182 tggaggtgtgcaagaggttctggtcggtcagtgaggaatcgtaccccgacacagcc 1241
 Db 1287 CGATATCATCTGACATGACCAATGCCCTGGTGGCAGACGCCCTCATCTTCACTCAAC 1346
 Oy 1242 gctcgaagaagaatcgtgtgacttcaagcagccggtgtgagacatgaag---aaatgag 1298
 Db 1347 CCGTTCACGCCCGTGACGCTCTGTGTCCAGTCTCCATGAGCGCATCTTGTGCTCCAG 1406
 Oy 1299 caagcagaatcgtgtgagcaacttgaaccagaatgtgaagctcatalcggatcatalca 1358
 Db 1407 CAGCGAGAGGCTATTTGAAGGCTGAAGAAACAGAAAGATATTCGTCAACTCAATGA 1466
 Oy 1359 gacctggaaagaaagcttgcgaagccgagaaatccacaaggaacgaaagcggtcgt 1418
 Db 1467 GACCTGGAGAGAGACCTGCGGCGAGACAGAACGCCATCCGATGAGAGGAAAGCCTTGGT 1526
 Oy 1419 cgaagagctgggt-----atcagcatcgaaagaggttltgtgaccttaccatccaa 1472
 Db 1527 GCGTGAGATGGGTGTGGCCATGAGGAGATGCTGTACCTGGGTGTGTCTTCCCA 1586
 Oy 1473 agaaatgacacatcgtatcaacttgaagcgatgactccttccttgcgtgagttgttcta 1532
 Db 1587 AAGACTCTCTCATCTTGTCAACCTGAATGAGACCTTTGATGTCCGAATGCCCTCTTA 1646
 Oy 1533 caacatcaagcccggtgcgaagaggttgaagacgtcaacccaagatatacaagcgaag 1592
 Db 1647 CTACATCAAAAGATGGCGTCAACAGATGGGCAAGAGAAATGACAGAGGCGCAACAT 1706
 Oy 1593 tgcctgaacggttcgaagatcctgaagaacacgttgcgttgaagaatgtg----- 1644
 Db 1707 TGTCTGAGTGTGACCTTCAAAAGAGAGACCTCATCTTCCGAGTACCTCCGAG 1766
 Oy 1645 -----gacaagcgtgtgacacatcgtgcgaacggaagagtcgtcgtatgttgaag 1697
 Db 1767 AGCGGAGAAAGCTGTGTGATACCTGAGGCCCTGTGAAGAGCAGACATATGTCAACGG 1826
 Oy 1698 cgtgcgaatcgaaacactactcgtcgcgcagcggtcgaagatcctcgtgagatt 1757
 Db 1827 CAAGAAGATCACTGACCCAGCATCTTACGGTCAAGAAACCGCATATGAGGAAGAAG 1886
 Oy 1758 ccaacatlttgcattcaacacatccggaagaagcgtgtgagagc 1802
 Db 1887 CCACGTGTTCCGGTTAACCAACCCCTGAGAGGCCGCTCAGAGACG 1931

RESULT 4
 AB023656 5739 bp mRNA linear ROD 24-FEB-1999
 LOCUS AB023656 Mus musculus mRNA for KIF1B-beta, complete cds.
 DEFINITION Mus musculus mRNA for KIF1B-beta, complete cds.
 ACCESSION AB023656
 VERSION AB023656.1 GI:4512329
 KEYWORDS KIF1B-beta.
 SOURCE Mus musculus (strain:ICR mouse) 4 weeks CDNA to mRNA.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (sites)
 AUTHORS Nakagawa,T. and Hirokawa,N.
 TITLE Identification and characterization of a new kinesin superfamily
 KIF1B-beta
 JOURNAL Unpublished (1999)
 REFERENCE 2 (bases 1 to 5739)
 AUTHORS Nakagawa,T. and Hirokawa,N.
 TITLE Direct Submission
 JOURNAL Submitted (15-FEB-1999) Terunaga Nakagawa, Grad. School of
 Medicine, Univ. of Tokyo, Department of Cell Biology and Anatomy;
 7-3-1 Hongo Bunkyo, Tokyo 113-0033, Japan

Db 1664 NCACCTAAAGAACCCACACCTCTTTAACTTCATGAAAGAACCCATTAATGTCGAGTGT 1723
Oy 1525 ctgtctacaacatcaagcccgagacaaggttggaaacgtcaaccaagatacaca 1584
Db 1724 TTGCTCTATTACATCAAGATGGAATTACAGGGTTGGCCAGACAGATCTAGCGGCC 1783
Oy 1585 gcggaatctcgtcgaacggttcgaagatcctgaagaacactgtacgttgaanaatgt 1644
Db 1784 CAGGACATAGTCTGATGTCGGGCCACATTAAGAGACACTTGTCTTTCGAGATGAG 1843
Oy 1645 gacaacg-----ttgtgacatcgtgcgaacagagaaggtctgtcatg 1689
Db 1844 AGGACCAACACTGCTGAAGATTATGTGACCTTGAGCCCTGTGAGCGCTCAACAAACATAC 1903
Oy 1690 gtgaacgctgtcgaatcgaacgaactactcgtcccgagcggtacagaatcctcgt 1749
Db 1904 GTGAATGGCAGAGAGGTGCTCATCTCTTTCAGCTGCGCTCAGGAGAACCTTATCATCAG 1963
Oy 1750 ggcgaattccacatttcgatccatccatccggaagaagctcgtgcgaacgcaaga 1808
Db 1964 GGTAAACACATGTTTTCGTTTAAACACCTGAACACGCGAGACAGAAAGGAAAA 2022

RESULT 5
LOCUS MUSKIF1B 4862 bp mRNA linear ROD 04-FEB-1999
DEFINITION Mouse mRNA for kinesin-like protein (Kif1b), complete cds.
ACCESSION D17577
VERSION D17577.1 GI:407338
KEYWORDS Kif1b; kinesin-like protein.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Nangaku,M.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-1993) Masaomi Nangaku, University of Tokyo,
Anatomy and Cell Biology; 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan
(Tel:03-3812-2111(ex.3336), Fax:03-5689-4856)
REFERENCE
AUTHORS Nangaku,M., Sato-Yoshitake,R., Okada,Y., Noda,Y., Takemura,R.,
Yamazaki,H. and Hirokawa,N.
TITLE Kif1b, a novel microtubule plus end-directed monomeric motor
protein for transport of mitochondria
JOURNAL Cell 79 (7), 1209-1220 (1994)
MEDLINE 95094296
REFERENCE
AUTHORS 3 (sites)
Nomura,N., Nagase,T., Miyajima,N., Sazuka,T., Tanaka,A., Sato,S.,
Seki,N., Kawarabayashi,Y., Ishikawa,K. and Tabata,S.
TITLE Prediction of the coding sequences of unidentified human genes. II.
The coding sequences of 40 new genes (K1A0041-K1A0080) deduced by
analysis of cDNA clones from human cell line KG-1
JOURNAL DNA Res. 1 (5), 223-229 (1994)
MEDLINE 96051398
FEATURES
source
1. 4862
/organism="Mus musculus"
/db_xref="taxon:10090"
293..3745
/codon_start=1
/product="Kif1b"
/protein_id="BAA04503.1"
/db_xref="GI:407339"
/translation="MSGASVAVAVRPPNSRETSKESCIITOMGNSTSIINPNPK
EAPKSPFDGYMSTSPEDPCFASONVYNDIGEMILHAEGVNCVCAFGCGAC
KSTYMGQESQAVIITPOLCELEKINDONKEMKSEVSYEITERYRDLNP
KNGMLRVEHPLEGLVEDLSKLAVTSTTDIADMDGKARVAAATMNETSSSH
AVFTVTFQKODPEPTNLSTEVSKISIVDLGSRASSTGAKTRLKGAINKSLT
TLGVYISALAEYSKKKKRKTDFPYRDSVYTLMLRNLGNSSTAAVAAADINDYD
TLSTRLVDAKOIKCNVYNEDPNKATLVRLKEFTRLDLRAOGLDIDTSGS
LTPSPSGSINSOGLTSTVTSIOERIMSPGCEVTAIERKSEKITAETNWEKLR
KTEATRMEALAEKVAIRDRDIDGVFSFKPIPHLVNEDPLMSCLLYTKDG

ITRVGADAERRDQIVLSGAHIKEEHLFRSRSNTGVIYTLPEPERSYVNGRV
AHPVLRSGNRILIMKNNHVRFNHNEOARAEKREKTSATSEPDWTFPAQELLEKO
GIDMEKEMERLQEMELIYKKEEADLEQORLADSDSDSDSKSEMLIT
SLREKPLSKLOTIYKKGIPSSGKREIKMYOIIPORRLSKDSKWTISLKTIOAV
KEICEVPLANDRHSROETELAIYKMEELCMKRNKPNENDSRVARDVMDVYV
GDEKTEDMNVTEKGGTDVDDLVHIDKLEDTILQEVKKONNMDEIKVIRNMLMEK
VLPLIGSOEKQSGSGHKTEPLVAQANSVDNGVSGESGELKBERESQNLNQPAP
RGRRLMMROEQIRFKNLQOQETLQRLQRNPHRFIPEPNRPPRPSKSNHRNSW
SGSTHITTEDVILRIPIKDEGKEKNESEOEKVGRAASDVOSAMGTROSDHIOY
SKOHSNOOPPOLMRMSNLNNGOPKTTROCATATSSSELSHSHPTADLOTPOAKR
HIHOHROPICNTNGQYEGYSTASCCQATDTPSHCNQFTVPPRMRQPSAPNLAGR
ETTV.
BASE COUNT 1435 a 1064 c 1200 g 1163 t
ORIGIN

Query Match 17.8%; Score 419.4; DB 10; Length 4862;
Best local Similarity 56.8%; Pred. No. 3.9e-108;
Matches 953; Conservative 0; Mismatches 681; Indels 45; Gaps 8;
Oy 166 atggatggcccggaagcatttcgctgcagtcggtcgtatgtgcttccttgacaag---aat 222
Db 422 AAGGAAGACACAAAGTCTTCAGCTTGAATCTCTACTGTGTCACACACTGCGCTGAA 481
Oy 223 gctcccaactatgcgagacagaagaacctatccaagatccggaagtcgcccttcgcat 282
Db 482 GATCCCTGTTTTCATCGCAGGAATCGTGTACAAAGACATTGGAAGAAAGAAATGCTTTA 541
Oy 283 aatgcattcaagggtataacaattatcttcgtcctacggtacgaacggttcgggcaag 342
Db 542 CATGCTTTGAAGCTTATATATGTCGTATCTTTCCTATGCGCACTGTGCTGGGAAA 601
Oy 343 tctattcaatgatgatgagcattgcaaggaag-----catgsgctgattcccgagatttc 396
Db 602 TCCTTACACATGATGCGGAAACAGAGAGACCAAGCTTAATCATCCACAGTTGTGT 661
Oy 397 caggacatgttccggtatlaaagacatgcgaggaaggaagaacctcattcgaacgtc 456
Db 662 GAAGAACTCTTTGAGAAGATCAATGACAAATTTGATGAAGACAAA---TGCTTACTCTGTA 718
Oy 457 gaagttctgacttggaataattcaatgaacgaggtgcgagactctgtaatccgtcgaca 516
Db 719 GAGGTAGCTCAATGGAATTTACTGTGAGAGATGAGATGGAATTTACTGAAATCCCAAAAC 778
Oy 517 aaggggaaatccaaggtcccgagacacccgtcgacggccctcagctggagagacttgagg 576
Db 779 AAGGTAACCTTGGCTGTGCTGCTGAACACCCCTTCTTGACCTATGTGGAGATCTGTCC 838
Oy 577 aagctgtcgtgcgacatcattccaagaatcgaagaatcctcattgtagtgcgacaaagcc 636
Db 839 AAGCTGGCGGTACACCTCTCACTGACATGCTGACCTGATGATGATGCTGGAAACAAAGCC 898
Oy 637 agaaaggttgcgcgacacaaacatgaacgagacatccaagtcgatacccaagcgcgtctact 656
Db 899 AGGACTGTGGCACCTACCAATGAATGAAGACAAGCAGCGCTGTCTCACTGCTGTTACAC 958
Oy 697 ttgaccttgagcaaaagtgcatagtatgaagagaccaaatatgacacagagaaggttcg 756
Db 959 ATTGTCTTTTACCCAGAAAGAACAGATCCCTGACACGAACCTTCTACTGAGAAAGTCACT 1018
Oy 757 aagatcaatctgtatgatttgcggttctgaagcgagcaacgtccacggaagcattcgtga 816
Db 1019 AAATACAGCTTGGTGTGATCAGACAGAAAGTGAACGGGCTGATTCACACTGTGCAAAAGGA 1078
Oy 817 ggcgcgactgaaggaaggtgtgaagatcaacgcgtcactcttcgagccctcaagtcgtgatc 876
Db 1079 ACTCGATTAAAGGAAGGTGAAATATCAATTAAGTCCCTTCACTTGGGCAAAAGTCAAT 1138
Oy 877 ggcgcgtacggaataatgcgtcgcggaacaaagaagaagatacagttgaagcattcaccga 936
Db 1139 TCGGCTTGGCAGAAA---GCTGAGTAAAGAAAGAAAGAAAGAAAGTATTTATTCCTTATAGC 1195

OY	937	gatttcgactacgctgagcttcttgaagacatcccttggagagaaactcgatgaccgcat	996
Db	1136	gatttcgactacgctgagcttcttgaagaaatcttggagaaatctccgaactcgaaatg	1255
OY	997	attgcgcgcatttcgcctgctgatatataactttgaagagactctcagtaaccctcgatat	1056
Db	1256	gtttgcctgctctgagccacccagcatcattcagatgaaacttttgagcactcgagatnc	1315
OY	1057	ggagactctggagagcgaatcaagaacccagcagcagtgtaatgaagaccggaacgcg	1116
Db	1316	ggagatcgctgcaaaacaaatttaattgcaattgcgttatcgaatgagaccccgaaatgcgaag	1375
OY	1117	atgataccgcaatltgaagaggaactcgcgcacacttgaagagcaaaact-----ccagac	1170
Db	1376	ctggtttcgagagattgaagaggaagcagccagctcgaagacactcttctgctctcaggsa	1435
OY	1171	agtgtctgaagtgtagagtggtgtgcaggaaggttcttgcgcgcagcttgaagaaatcgtaaccg	1230
Db	1436	ctggagagatattatttgatatacatccatggggctccctcaccttctcattccatcattcagctca	1495
OY	1231	ccgcgacgcgcctcgaagagcaaaactcgtctgcatctcagcagcagctgacagctcaag	1290
Db	1496	ctttaattagcagagtgggccttaacatctgtgaccacagatcca---ggagaggaatcatgtct	1552
OY	1291	aaatgagcaaggcagaaatcgttgaagcaactgaaccagagtgtagaagctcatacggat	1350
Db	1553	acacctggagggaggaagaaacccattgaacctcgaagagatcagacagacatcattgcctgac	1612
OY	1351	ctcaactcagaccttggagaagaagaagcttgcgcaagacgcggaagaatcacaggaacgagaa	1410
Db	1613	tttgatgaaacctgggaamaagaaacttctgtaaaacagagcccatcagaatggaagagaa	1672
OY	1411	gcgcgcctcgaagagctgggtatcagcat-----cgaaaagggtcttgttgcgcttac	1464
Db	1673	gccttggctggcctgagatggaggtggcattacacggaagatcgggggacatttgagagtttcc	1732
OY	1465	caatcctaaagaaatgccaatctagtcaacttgaagcagatgataccctcttctgctgaagt	1524
Db	1733	tcaccttaaaaaacccacacattctgttaaccctcaatgaagacccattgaatgctggagttct	1792
OY	1525	cttgctcaacaactcaagccgcggcagacaagaagtggttgaagacgtcaacaaatatacaaa	1584
Db	1793	ttgcctctatttattatcaagatggaatttacaaggtttggccacagcagatcttgaagcggcc	1852
OY	1585	gcgcgaatctgctgaacgcgttcogaagatcctgaagaacactgtacgttgaaaaatgty	1644
Db	1853	caggacattagtcctagctggggcccaacattaaagagagacattgtctcttggaggtgag	1912
OY	1645	gacacag-----ttgtaaccatcgttgcacaaagagaagcgtctgcatg	1689
Db	1913	aggagcacacactcggtgaagcttattgtgacctttggagccctgtgagcgcctcagaaacatnc	1972
OY	1690	gttgaacgcgttgcgaatcgacaagactactcgcctcgcgaagcgtgtacagatacgtctg	1749
Db	1973	gtgaatggcamaagaggtgctcatctctgttccagctgcctcagggaaacggtatcatcang	2032
OY	1750	ggcgcattccacattttctgattcaacatcccggaagaagcttcgtgcgaaacgcgcaaga	1808
Db	2033	gcttaaaacacattgtttctgttttaaacaccccttgaaacaagcccgagcaaaaggagaaa	2091

RESULT	6
AY043362	
LOCUS	AY043362 5575 bp mRNA linear PRI 18-AUG-2001
DEFINITION	Homo sapiens kinesin superfamily protein 1B (KIF1B) mRNA, complete cds.
ACCESSION	AY043362
VERSION	AY043362.1 GI:15212239
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE TITLE	JOURNAL	AUTHORS	TITLE	JOURNAL	FEATURES	source
1 (bases 1 to 5575)		Park, M., Shin, H., Lee, Y. M., Moon, E., Choi, W. and Kim, W.	Identification of the human ortholog of mouse Kif1b, a kinesin superfamily motor protein	Unpublished		
2 (bases 1 to 5575)		Park, M., Shin, H., Lee, Y. M., Moon, E., Choi, W. and Kim, W.	Direct Submission			
Submitted (03-JUL-2001)		Submitted (03-JUL-2001)	Biological Science, Ajou University, San 5 Wonchon-dong, Paldal-Gu, Suwon, Kyonggi-do 442-749, Korea			
Location/Qualifiers						
1..5575			/organism="Homo sapiens"			
/db.xref="taxon:9606"			/chromosome="1"			
/map="1p36.22"			1..5575			
/gene="KIF1B"			5..3466			
/gene="KIF1B"			/note="similar to mouse KIF1B; similar to KIAA0591 and KIAA1448 sequences found in GenBank Accession Numbers AB011163 and AB040881, respectively"			
/codon.start=1			/product="Kinesin superfamily protein 1b"			
/protein.id="NAK85155.1"			/db.xref="GI:15212240"			
/translation="MSGASVAVAVRVPFNSRETSKESKCIIOGNSSTIIPNPKP EAPSFEDSYWSTSPEDPCFASORVYNDIGKEMLLAFEGYANVCIFAYGOTGAG KSYTMGKQESQAGITPOLCELFEPNNHNGNENMSYSVEVSMETICEYRDLNLP KKNGLNVRDHPLEPVEYEDSLKAVTSYITDIADLMGNAKRVAAATNMETISRSIS VYIVTVEYQKHNETLIVATVYKSTSLVDIADMGDRASTGAGTRLKEGAIIRKSLTL TLGVLSIALAEVSKKKRKTDFIPYRDSVLTWMLRENGNSRTAMVALSPADINYDE TLSTLRADRAKQIKCNAVINEDNPNAKIVELKEVYRNLKDLTAOGIGDIIDTSMGSS LTSPSSCSLSOYGLTSTVTSIOERIMSTPGGEALIRLRESEKIIIEINMEWEKIR KTAIRNREKRLIAEMGVATIREDDGTLGVTSRPTPIVNLNEDPLMSECLITYIKRG ITRGQDAERKODIVLSGAHIKECHTFPNSERNSEVITLLEPCRSEYIVYGVKKRV GIDPQVLRSGNRIIMGKNHVERFNHPEDQARERETPEAETPSEVDWTFQORELLEKQ SGPOLRSGMERKILCEMELIYKKEEADLLLEOQLDADSDGSDSKRSEESWKILTEQ SLRKLRLPSPKLOTIVKKGCLPSSGSKREPIKIMOIPORRLSKDSKWTIISDLKIOAV KETICEVAINDFRHSROETELAIIVKKMKICAMYGKDPNDRSDRAVARVMDYTGCV GDEITIEVMAATGKSTIVYDILKIHIDLIDLEIYKKNKKMDEIYVLRKMLKMK VLPILTSQEQSPGSHNAKEVGVGVSSTSENNVSKDNCLELAEEVSLDMNDDPAF RRGRLRMWRQEQIRFKNLQOQETIKYLRQNVNPRFTPEPNKRKRFPPKSPKRNISM SPGHIIILITEDEVLEIRLIPKDDARKNGKESQGGKAFKDPQPMGSGSDQSDH IOVGHKIIINNOQEPOLRMRNSNSLNNGOPSTRCQASASAESLNSHSGHPADVQTFQ AKRIIHOHROSICYNNTNGQGLEGNAAISYQKQDKPCHSCQFVTPPMRRQFSAPLK AGRRTTY"						
26..1072			/gene="KIF1B"			
/note="Region: KISc ATPase domain; catalytic domain of motor"			35..1120			
/gene="KIF1B"			/note="Region: motor domain"			
1502..1750			/gene="KIF1B"			
/note="Region: FHA domain"			1101 c			
1283 g			1471 t			
BASE COUNT			1720 a			
ORIGIN			1101 c			
1283 g			1471 t			
Query Match			16.9%: Score 398.4; DB 9; Length 5575;			
Best Local Similarity			55.9%: Pred. No. 4e-102;			
Matches 941; Conservative			0; Mismatches 656; Indels 45; Gaps 8;			
157 aaactatattgtagtgcgcgaagggcatttcgcttcgatacgtcgtatggtccttgac			216			
125 AAGATCCAAAGAAAGCGTCCAAAGTCCTTCGACATFTCTACTGGCTCATGCC			184			
217 ---aagaaatgcccaactatgcgagacaggaagaacctatccaaagtctcggagtc			273			
185 TCACCCGGAAGATCCGTGTTTTCATCTCGAAGACCGTGTGTCAATAGACATGGCAAGAA			244			

Oy	274	ctctcgtataatcatcattcaagggttatbaacaattgatctcttcgcctaagtcgaacggt	333
Oy	275	ctctcgtataatcatcattcaagggttatbaacaattgatctcttcgcctaagtcgaacggt	334
Db	245	ATGCTCTTACACGGCTTTGAGGGATTAATAGTCTGTATTTTGGCCCTATGGGACAGCTGT	304
Oy	334	tcggcgaagtcctcattcaatgatctggcgctatgagcaag-----agatatggcgtgatccgg	387
Db	305	GCTGGAAAATCTTATACAAATGATGCGTAACCAAGAAAGAACCGAGCTGGCATCATTTCCA	364
Oy	388	cgaatttcgcaagacatgcttcgcgcygaltatgaactcgagaactgagaaggaacaagactact	447
Db	365	CAGTTATGTGAAAGAACTTTTGGAGACAAATCATCAGACAACGTAATGAAGAAA---TGTC	421
Oy	448	tgcagcgtcgaagtttcgctactcttggaaatttaacaaagaaacgagltggaaccttcgcat	507
Db	422	TACTCTGTAGAGGGAGACTCATATGGAATTTACTGTGAAGATATCAGAAATTTGGCTGAT	481
Oy	508	ccgtctgcacaaaggggaaattcccaaggtccgaagacacccgtcgaacccgcctccacgttgag	567
Db	482	CCAAAAAACAAGGGTAAATTTGGCGTGCGGATACCCCACTTGCTGGACCCATATGGAG	541
Oy	568	gacttgcgcaagctcgtgctcgtcgatcatctccaagaaatcgaaaatcattcatgatgagggc	627
Db	542	GATCTTTCACAAAGTGGCAGATTACTTCTCTACACAGACATTCGTCACCTCATATGATGCTGG	601
Oy	628	aacaagaccgaagacggtctgcgcgcacaaatgaacgaagacatccagtcgactccacgcgc	687
Db	602	AACAAAGCCAGGACAGTGGCGACGTACAAATGATGAATGAACAAAGTGGCCGTTCCACGTT	661
Oy	668	gtcttcacatttgaccttgcagcacaaggtgycatgatgaagagaaccaaaatgycacaaag	747
Db	662	GTTATACGATTTGCTTCCACCACCAAAACAGATATATGAGACCAATCCCTTGCCACTGTG	721
Oy	748	aaggttcggaagatcagttcgtgatatttggcgggttcttgagagacaacgttccaccga	807
Db	722	AAGGTATGATTAACCACTTGCTGTGATCTAGCAGGAAGTACCGACCTGCTTAACTGCT	781
Oy	808	gctactggaagcgacacgaaagaggtgycagagaatcaacccgtcacttctcgaacctcagtt	867
Db	782	GCCAAAGGACCTCGCTTAAAGGAAGAGACATCATTAATATGCTCTTCAACTTTGGGC	841
Oy	868	ctgtgtgtatgagcgcagcgcgatatgtctgcgggaaacagaagaagaatcagttatga	927
Db	842	AAAGTCATTTTAGGCTTTGGCCGA---GCTAGTAAAGAAAGAAAGAAACAGATTTTAT	898
Oy	928	ccttaaccgagatctcgtcattcagcttggtctctcgaagcactccttggagagcaactcgatg	987
Db	899	CCCTACAGGAGTTCTGTATCTTACTTGGCTCTTCGAGAAATTTAGTGGCAATTTCCGG	958
Oy	988	accgcacatgattcgcgcacattcgcctgtcgtatataacttgaagagactcagttacc	1047
Db	959	ACTGCAATGGTTGCTGCTGTCAGCCCCCGGCGATCATCACTACATCATAGACTTTGACACT	1011
Oy	1048	cttgcataitgagcactcttcgcaagcgaataaagaacaaacgacgttgttcaatgaagccgg	1107
Db	1019	CTGAGATATGCAAGATCGTGCAAAACAAATTAATATGCATCTGTTTATCAATGCGACCCC	1077
Oy	1108	aacgcgcgagatgcatcgcgcgcgcaglttgaagaggaactcgcagctctgagagcaaac	1166
Db	1079	AATGGCAAACTGTTTCGGAATTTAAAGAGAGAGTGACACGGCTGGAAGGACCTTCTTC	1133
Oy	1164	---ccagagcagttgttgaaggttgaaggttgcgtgcaggaaggtcttcgcygggcacagtgaagaa	1222
Db	1139	GCTCAGGGCCGGAAGATTAATTAATGATACATCCATGGGGTCCCTCATCTTATGCCCATCT	1199
Oy	1222	tcgtaccgcgcgcgaacgcgcgcctcgtagaagcaaaatcgtgtcgtattcaagcgcgcgatgcg	1287
Db	1199	TCTCTCTACATCAGATGACAGTGGGCTTGACGCTCTGTGACAGATTTTCA---AGAGAGG	1255
Oy	1282	acagtcacaagaaatgaagcagaagacaatcgttgagacacttgaaccaagatgagaagctc	1344
Db	1256	ATCATGCTCTACACCTGGAGAGAGAGAACTATTGCAACGTTTAAAGAAATTCAGGAAGATC	1311

QY	1342	tatcggtctcaatcagagccctgggaagaagagcgcgcgaagcagcgaggaatccacag	1401
QY	1342	tatcggtctcaatcagagccctgggaagaagagcgcgcgaagcagcgaggaatccacag	1401
Db	1316	ATTGCTGAGTTAAAGAACTTGGAGAGGAAGCTTCTGTAATAACAGAGCCATCGAATG	1375
QY	1402	gaacagaaagcgcgcctcgaagagctgggtatcagcat-----cgaaaaggcttllgt	1455
Db	1376	GAGAGAGAGCGCTTTGTGGCTGAGATGGAGATTGCCATTCCGGGAAGATGGAGAACCTTA	1435
QY	1456	ggcccttaccatccaaagaatcgcaatctagtcacacttgagcgatgatccctctctg	1515
Db	1436	GGGGTTTCTCACCATAAAAGACCCACATCTGTTAACTCTCAATGAAGACCCACTAATG	1495
QY	1516	gctgagtgctgtcttacaacatcaagcccgcgagcagaaggttggaagcgtcaacca	1575
Db	1496	TCTGAGTGGCTACTTATTATACATCAAGAATGAATTACAAGGTTGGCCAGACAGATCT	1555
QY	1576	gatacaaaagcgaatctcgtctgaacggttcgaagatccttgaagaacactgtacgtt	1635
Db	1556	GAGCGGCGCCACGACATAGTGTCTGAGCGGGCGCTCACATTAAGAAGACCATTTGTCTTC	1615
QY	1636	gaaatgtggaacaaag-----tttgacacatcgtgtgcaaacgaagaagct	1680
Db	1616	CCGAGTGAAGAGCAACACGCGGGAGATTATCGTAGACCTTAGACCCCTGTGAGGCTCA	1675
QY	1681	gctgtcatgttgaacgcgctgcgaatcgcgaacacactcgcctcgcgcgcgcgcacag	1740
Db	1676	GAAACCTAGTAAATAGGCAAGAGGGGTGTCACGCTGTTCACACTGCGTCAGGAACGCT	1735
QY	1741	atcatcctggcgcatctccacatcttttcgatccaacatccgcgaagaagctcgtgcgaa	1800
Db	1736	ATCATCATGTCGTAATAAACCATGTGTTCCGCTTTAACCAACCCCGAACAAGCACGACCTGAG	1795
QY	1801	cg 1802	
Db	1796	CG 1797	
RESULT	7		
AF257176		6816 bp	mRNA linear
LOCUS	AF257176		PRI 02-MAY-2001
DEFINITION	Human sapiens kinesin superfamily protein KIF1B mRNA, complete cds.		
ACCESSION	AF257176		
VERSION	AF257176.1	GI:13925306	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 6816) Yang,H.W., Takita,J., Chen,Y.Z., Soeda,E., Piao,H.Y., Hashizume,K. and Hayashi,Y.		
TITLE	Cloning of human kif1b gene that maps at 1p36, which is homoygously deleted in neuroblastoma cell line NB1 and mutated in some other neuroblastoma cell lines		
JOURNAL	Unpublished		
AUTHORS	2 (bases 1 to 6816) Yang,H.W., Chen,Y.Z., Takita,J., Soeda,E., Piao,H.Y. and Hayashi,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-Apr-2000) Pediatrics, The University of Tokyo, Hongo		
FEATURES	7-3-1, Bunkyo-Ku, Tokyo 113-8655, Japan		
source	Location/Qualifiers		
	1..6816		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="1"		
	/map="1p36.2"		
	154..5466		
	/note="kinesin related protein"		
	/codon_start=1		
	/product="kinesin superfamily protein KIF1B"		
	/protein_id="AAK49332.1"		
	/db_xref="GI:13925307"		
	/translation="MSGASVVAVAVRPNSRETSKESKCIOMQGNSTSIINKNPK"		
CDS			

Db 1157 CCGGTGCATGCCATCATCAACGAGAGCCCTTATGCGCGGTGATTAGAGACTGCAGGA 1216
 QY 1137 ggaactgcgcagctggagggcaaac-----ccagagcaagtggt 1176
 Db 1217 GGAAGTACCCGGCTGGGGAACTGCTGATGCTCAGGAGCTGCAGAGCTCTGCTGGA 1276
 QY 1177 ggaagtggaagtggtgaggaaggttct-----gagcgcaagtggaaggaactgtaacc 1229
 Db 1277 AGGCTCAAGACGAGAGAGGAGGATGTCAGAGGCGCCCTGCACACTGTCATCTCCCC 1336
 QY 1230 gccgcgaacgcgcgtcgaagaagcaatcgtgtcgtatcagcagccgagatgagcaagtc-- 1287
 Db 1337 AGCTCAGTTTACCCCTCATCCACCACACATTAATGGGAGCTGGAGCGCTCATTTCTC 1396
 QY 1288 -----aagaaatgagcaagcgcaagaatcgttgagcaactggaacccagagtga 1334
 Db 1397 CCCCACACGAGAGTCCAGATTTGGGCTGAGGAAGCCATGGAGAGGCTGCAGAGACAGA 1456
 QY 1335 gaagctcattcgagatcattcaatcagacctgggaagagaagcttgccaagaccgagaaat 1394
 Db 1457 GAAGATTATAGCTGAGCTGAACGAGACATGGAGAGAGACTGACGAAGACAGAACCCCT 1516
 QY 1395 caacaaggaagcaagcgcgctcgaagagctggtatcagcgc-----gaaaggg 1448
 Db 1517 GAGGATGAGAGAGAGACATTTGCTGAGATGGGGTGGCGCTCCGGAGAGATGGGGG 1576
 QY 1449 ctgtgtggcccttaccacacccaagaagaatgccaatcagatcagcttgaagcagtgatcc 1508
 Db 1577 AACGTGGGCGCTTCTCTCTCCAAAGAAAGACTCCCACTGGTGAACCTGAACGAGACCC 1636
 QY 1509 tctctggtgagtgcttctgtcttacaacatcaacacccgagcagaagagtggtgaaact 1568
 Db 1637 TCTGATCTGAGTGTGCTGCTACACATCAAGATGGCGTCAACAGAGGCTGGCGAAGT 1696
 QY 1569 caaccaagatacaacaagcggaatctgtctgaagcgttcgaagatc---ctgaaagaaca 1625
 Db 1697 AGATATGACATCAAGCTGACCGGACAGTTTATTCGGAGAACACTGCTGTTCCGGAG 1756
 QY 1626 ctgtgactttgaaatgtgacaacgtgtgacacatcgtgccaagcagaagcgtgctgt 1685
 Db 1757 CATGCCCAAGCAGATGAGAGAGTGTGTGTCATCTGAGCCTTGTGAAGAGCTGAGAC 1816
 QY 1686 catgtgtgaacggcgctggaatcgaacgaagcactacgcctccgacgagcgttacaagatcat 1745
 Db 1817 ATATGATTAATGGGAAGCTTGTGACGAGCCGCTGAGTGAAGTACAGGAATAGATGT 1876
 QY 1746 ccgtgaggaattcgaacatttgcattcaaccatccggaagaagcgtgtgaggaagcga 1805
 Db 1877 GATGGGCAAGAACACGTTTCCGCTTCACCCCGAGCGAAGCGTGAACGGGA 1936
 QY 1806 a 1806
 Db 1937 A 1937
 RESULT 9
 AB014606 6217 bp mRNA linear PRI 06-FEB-1999
 LOCUS Homo sapiens mRNA for KIAA0706 protein, complete cds.
 DEFINITION AB014606
 VERSION AB014606.1 GI:3327225
 KEYWORDS Homo sapiens adult male brain cDNA to mRNA, clone_1ib:pbluescriptII
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 6217)
 AUTHORS Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
 TITLE Direct Submission
 JOURNAL Submitted (26-MAY-1998) Osamu Ohara, Kazusa DNA Research Institute,
 Laboratory of DNA Technology, Yana 1532-3, Kisarazu, Chiba
 252-0812, Japan (E-mail:cdna@nifekazusa.or.jp, Tel:+81-438-52-3913,
 Fax:+81-438-52-3914)

REFERENCE 2 (sites)
 AUTHORS Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A.,
 Kotani,H., Nomura,N. and Ohara,O.
 TITLE Prediction of the coding sequences of unidentified human genes. X.
 The complete sequences of 100 new cDNA clones from brain which can
 code for large proteins in vitro
 JOURNAL DNA Res. 5 (3), 169-176 (1998)
 MEDLINE 98403880
 FEATURES
 SOURCE
 location/Qualifiers
 1..6217
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HG04286"
 /sex="male"
 /tissue_type="brain"
 /clone_1ib="pbluescriptII SK plus"
 /dev_stage="adult"
 368..3679
 /gene="KIAA0706"
 368..3679
 /gene="KIAA0706"
 /note="HG04286 cDNA clone for KIAA0706 has a 23-bp
 insertion at position 2033 of the sequence of KIAA0706."
 /codon_start=1
 /product="KIAA0706 protein"
 /protein_id="PAA31681.1"
 /db_xref="GI:3327226"
 /translation="MAGASVAVAVRPVPPNARETSQDAKCVSMQNTTSITINPKSG
 DAPKSTFEDSYMSHTSTEDPEFASQOQVVRDIEEMILHAFEGYNVICIFAYGGTGG
 KSYTMGROREGOQIVPOLCEDLEFSEVSEASOLSSEVSEMEYCEVRDLNP
 KRSGLRMRREHPIGLPYODISKLAVSYADIADIMCCGKARVATNNMTSSRH
 AVETIVTFORCHDITGLDSEKVKISLVLDAGSERDSSGARBRKBEKGINIKSLT
 TIGKIVTIALDMOSKKRKRSPYRDSVLTKMLKENGCSRTMTAISPADINVEE
 TSLTRADRTKQTRCAITINEDPNALIRELOEYVRLRLMLAGLSASALEGLT
 EEGSVRGALPAVSSPPAPVSPSTTNGELESEFSNTSQTGPEPAMRLQETKRI
 IALENETMEERLKRTEALRMERELALMEGVAVVEDGTGVEFSPKTPHLYINEDP
 LMESECLLYHKIDGTVRVGVDMDKITGQFIREQHCLFRSIPQDGVVTVLECEGA
 ETVYNGKLTPEPLVKSGNRYVMGKNVFRNKHEDLLEORLVLADSDGSDSK
 NFAOKELLBOGQIDIKLEMRRLQDLGNOYRKEEADLLLEORLVLADSDGSDSK
 RSCRESRLISLEBOJPRPTVOTIVRCGIPSGKRRAPRYQIQRRLQSKDR
 WATMADLMQVKEICTEVALDRHRAETELALUKMEELCTYKTPGPGDAMPA
 VARDVPTVEEESGGAGSGGSEEGARAEVEDLRHIDKLTGILQEVKIQNSKDR
 ELQALRDRLMERVPIPLADHEDENEGGEVPAAPPEGSAAEAPASDMPARPP
 SPPLSWEVRSRLNEEDPAFRGRRLRYLKKOBLQLOGSGGGGLRRPRAFEVPP
 HDCKLRPPKSNPQRESPWGMGSGEAPTLQDPEEYTPHPATPARPPSRSHHNR
 RNSIDGGRRRGASAPREPHNRPKHNSTPPRPYRQRPGRPRPYRTTPRRAR
 RQNSAPDLKESGAAY"
 BASE COUNT 1406 a 1723 c 1892 g 1196 t
 ORIGIN
 Query Match 16.2%; Score 382.2; DB 9; Length 6217;
 Best Local Similarity 55.0%; Pred. No. 1.7e-97;
 Matches 1024; Conservative 0; Mismatches 738; Indels 99; Gaps 10;
 QY 6 gggcggtggaatatcaagtggtgtggtgagccgttccagccgcaagaatcga 65
 Db 370 GGCTGTGCTCCCTCGGGAAGTGGCAGTGAGGCTTGCCCTTTAAAGCCCGGAGACCAAG 429
 QY 66 ccgttgcgcaaatgtattgttgcgagtaggaagaaatacaacacatccaccctccctcc 125
 Db 430 CCAGATGCCAAGTGTGTGTCAGCATGACAGGGCAACACACCTTCATCAATCCCT-- 487
 QY 126 ggggtgcgaagaagaagcgcgtaaaagtgcgaacactttatggtggtccggaagcatt 185
 Db 488 -----AAACAGAGCAAGGATGTCGCCCAAAAGCTT 516
 QY 186 tgcgttcagatcggtcgtattgtc---cttggacaagaatctcccaactatgcgagaca 242
 Db 517 CACCTTTACACTCTCTACTGCTACACACTTTCGACGAGAGACCCCGATTGGCATCTCA 576
 QY 243 ggaagaacctattccaagaatctcgagctccgcttccttgataatgcatccaaggttataa 302

```
Db 577 GGAGCAAGGTATCGGACATTGGAGAAAGATGCTGCTCAAGCCCTTTGAAGGCTACAA 636
Oy 303 caattatcatctcgctacgctcagacacggttcgcggaagctctatcatatgatgg--- 359
Db 637 CGTGTGTCATCTTTGCTTATGGGACAGACCGGGCTGGGAAATCCTATACCATGATGGGGCG 696
Oy 360 ---ctatggcaagagacatgagcgtlgaatcccgagatltgcagagacatgtctcggcgtat 416
Db 697 ACAGGAGCCAGGAGGACAGGAGGATCGTGGCCAGCTCTGTGAGGAGCATCTTCTCTGGCGT 756
Oy 417 taatgaactgcgaagagacagaacacctaacttcgacgcgtcgaaglttcgtaacttgaat 476
Db 757 TAGTGGAAACAGAGAGTGTCTAG---CTATCTACTCTGTGGAGGTGAGCTATATGAGAGAT 813
Oy 477 ttaacaatgacgagtgacgagacttcgtgaatccgctcgaacaaagggaactcaagatcgc 536
Db 814 CTACTGTGAGCGGGTACGAGACCTCTTGAAACCCAAAGATCGGGGTTCTCTCGGGGCTCG 873
Oy 537 aagaacacccgctcgaacccgctcagctgagagacttgcgaaagctgtgtcgtcgaatcat 596
Db 874 GGAGCAAGCCCATCTCTGGGCGCCGACGTGACAGACTGTCCAATTTGGCTGTGACCTCTTA 933
Oy 597 ccaagaatcgaataatctcatgtagagggcaacaagccagaacggttcgcgcgaaca 656
Db 934 CGCAGACATTTGCTGACCTCATGACTGTGAAATTAAGCAGGACTGTGCTGCCACCA 993
Oy 657 catgaacgagacatccagctcgaatcccaacgctcttcacacttgaacttcgacgaagaat 716
Db 994 CATGAAATGAGACACGACGCCCTTCCATGCGCTTTACCATCTGTTTACACACAGCGCTG 1053
Oy 717 gcatgataagaagacccaataatgacacagagaagltgcgaagatcagctgtgtagatlt 776
Db 1054 CATGACACGCTCACCGGGCTGAGCTCGAGAGAAGTCAAGTCAAGTCAAGTGGTGACCT 1113
Oy 777 ggcgggttctgagcgagacacgtccacgcggaactactgtgaagcgcaactgaagaggtgc 836
Db 1114 TGTGGGAGTGAAGCGAGCGGACCTCTCAAGGGGCCCGGCGCATGCGCTTCAAGGAAAGAGC 1173
Oy 837 agagatcaacgcgtcaacttcgacccctagtgctgtgatatgcgagctagcgagatgac 896
Db 1174 CAACATCAATTAAGTCCCTACTACTACATAGGAAAGTGAATGATCGGCCCTTGACAGATTTGA 1233
Oy 897 gtcggaagaaacagaagaatcagltagtlaccttacacagatltcgatcgaagcgtgc 956
Db 1234 ATC---AAGAAGCGAAAGTCGGAATTTATCCCTTACAGGAGACTGTGTCTACCTGCGGT 1290
Oy 957 tctgaagacgtccttgggagggcaactcgaatcgatgcgaatgtccgcaatttcgcctgc 1016
Db 1291 GGTCAAGGAAATTTGGGGGGAACTCAGCACAGGACCATGATTGACGCCCTGAGCCCTCG 1350
Oy 1017 tgatatatacttgaagagactctcagtaaccttcgatatgaggaacttcgagagcgaat 1076
Db 1351 TGACATCAATTAAGGAGAGACTCTCAGCAACCTCAGAGTATGCTGACGACGCAACGAAGAA 1410
Oy 1077 caagaacacgagtggtlcaatgaagaccggaacgcgcgagatgatacgcgaglttgaaga 1136
Db 1411 CCGCTGCAATGCCATCATCAAGAGAGACCTTAATGCCGCGTGAATTAAGAGAGCTGACAGA 1470
Oy 1137 ggaactcgcgacgctgagagagcaaat-----cgaagcgaatggt 1176
Db 1471 GGAAGTAGCCCGGCTGCGGGAACCTGATGCTCAGGACCTGTCAGGACCTCTGCTGTGA 1530
Oy 1177 ggaagtgtgagtgctgagagaggtct-----ggcgggacgagtgagaatcgtlacc 1229
Db 1531 AAGGCTGAAGAGGAGAGAGAGAGTGTACAGAGCGCCCTGCGCAGCTGTGTCTATCTCCCC 1590
Oy 1230 gcccgacacgcgcgtcgcagagaagcaatcgtlgtcatltagcagcgagatgcgaagcagtc-- 1287
Db 1591 AGCTCCAGTTTCAACCTCTATCACCCACACACATTAATGGGAGCTGAGACCGCTCATTTCTC 1650
Oy 1288 -----aagaataatgacgaagcagaatacgtgtggagcaactgaaccagagtga 1334
```

```
Db 1651 CCCCAACAGGAGTCCCAAGATTGGCCTGAGAGAAAGCCATGAGAGGCTCCAGACACAGA 1710
Oy 1335 gaagctctatcggatctcaatcagaccctggaagaagcgtgcgaagacaggagaat 1394
Db 1711 GAAGATTATAGCTGAGACTGACAGACATGGAGAGAAAGCTTACGACACAGAACCCCT 1770
Oy 1395 ccacaagaacagaaagcgcgcgtcgaagagctggtlatacagatc-----gaaaggg 1448
Db 1771 GAGGATGAGAGAGAGAGCAATTGCTGCTGAGATGGGGGTGGCCGTCCGGAGAGATGGGG 1830
Oy 1449 cttgtgtgccttaccactccaagaatgccaatctcagtaactgaatgaatgatacc 1508
Db 1831 AACTGTGGCGTCTTCTCTCCAAAGAACTCCCACTGCTGTAACTGAACGAAACGCC 1890
Oy 1509 tcttcggtcgtgagtgctgtgtctacacatcaagcccggaagcagagaaggttggaaagt 1568
Db 1891 TCGATCTCTGAGTGTCTGCTCTACACATCAAAAGATGGCGTACACAGGCTGGCCAAAT 1950
Oy 1569 caaccaagatacaacagcgaatcgtctgaacggttcgaagatc---ctgaagaaga 1625
Db 1951 AGATATGACATCAAGCTGACCGGACAGTTTCATTCGGAGCAACACTGTCTGTTCCGGAG 2010
Oy 1626 ctgtacgttgaataatgtgagacagctgtgtaccatcgtgcgaagagagagctcgt 1685
Db 2011 CATCCCAAGCCAGATGAGAAAGTGTGTCTGAGACCTTGTGAAGGACCTGAGAC 2070
Oy 1686 catgtcgaacgcgtgtgagatcgaacagcctactgcctccgcgagcgctacagatcat 1745
Db 2071 ATATGTGAATGGGAACCTTGTACGAGAGCCGCTGCTGCTGAATCAAGGAAATGGAATGT 2130
Oy 1746 ccttggcgatltccacatlttcgatlcaacacalcgcggaagagcgtcgtgcgaagcga 1805
Db 2131 GATGGGCAAGAACACGATTTCGCTTCAACCCAGGAGACGAAAGCTGTGAACGGGA 2190
Oy 1806 a 1806
Db 2191 A 2191

RESULT 10
RNAJ696
LOCUS RNAJ696 4117 bp mRNA linear ROD 01-SEP-1997
DEFINITION Rattus norvegicus mRNA for a novel kinesin-related protein, KIFD.
ACCESSION AJ000696
VERSION AJ000696.1 GI:2370434
KEYWORDS kinesin-related protein.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Granulata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 4117)
AUTHORS Rogers, K.R.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-1997) Rogers K.R., Precinical Veterinary
Sciences, University of Edinburgh, Summerhall, Edinburgh., EH9 1QH,
SCOTLAND
REFERENCE 2 (bases 1 to 4117)
AUTHORS Rogers, K.R., Griffin, M. and Brophy, P.J.
TITLE The secretory epithelial cells of the choroid plexus employ a novel
kinesin-related protein
JOURNAL Unpublished
FEATURES
location/Qualifiers
source 1..4117
/organism="Rattus norvegicus"
/strain="Wistar"
/db_xref="taxon:10116"
260..3553
/gene="KIFD"
260..3553
/gene="KIFD"
/codon_start=1
/product="kinesin-related protein"
```

BASE COUNT 956 a 1101 c 1269 g 791 t

ORIGIN

/Protein_id="CAA0248.1"
/db_xref="GI:2370435"
/db_xref="SPTREMBL:O35787"
/translation="MAGSAVVAVAVKVRFPNARETSQDARKVSMQGNNTSIINPKOSH
MFLKSFDFSTWMSHTSVEDPQFAQQOQYTRDIGEMMLHAEGYVVCIFAAGTQAG
SYTMNGRQPGQOQIVPOLCEDELSRVANNSAQLSYEVSMTIYERHDLINP
SGSLRVREHPIILGPPYVDLSKLAVTYADLADMDCKNRVTAAATMNTSSRSH
VFTVYFTRSHDQHLGIDSEKYSKISLVNLGSEBDSGARGMLKGCANINSTIYH
LGKYSALADLOSKRRKDSFPIYRDSVTYMLKLEGNSTAMTALDSPADINYEEL
LSTAYADRTKQRCNATINEDPNARKIARELOEYARIRRELMAGSLASDGLKAVR
EGSGGVTPAASSPPAPASPPSPPPHNGELSEPSASPEQIPHLVINENEDPLMSK
ABLNRTWEKLEKTBALMEREDALLAENGSGQWTVSPKTPHLVINENEDPLMSK
BGLVLIHKGVTRVGQVDVLIKLGQFIREHOCLERSIPQDGVGEWMLBPCBGAETL
VNGKILVTEPLVAKSGNRIYMGKNHVFNRHNEDEQALREBERGVPPPGVPVWNF
ESDWLEQOQIDIKLEMERKLODLENOYKREKEEDLLEOQRLVADSGSDPKRS
KESMTLISLDELPEPTVQTVYKCGIAPSGSKRRAPRAYQIPORRLOQDKRWMT
MADLMQAVKEICEVALADVRHGAETELALAKMRKLTCTYKCPBPGAMAVAVR
DVMDVPEEGCGGGGGGGGGERGARGAEVEDLRAHDKITGLIOYKQKNSDKRELOA
LRDMLRMERVIPLTODLEDNEESGLVTYAPPEBSRFAVEAASNDSPAYRSPSP
SMSEVYSRLMEDDPAFRKGRIRWMLQOBDLQGLQSSGKRGGLRRPARKVPHDCK
LRFPPKSNQHRSESPMGSGSEAPQPEEVTAPPPGNRRPSPSPRPHPRRNNSLDL
GGSSRGSGSTQPEQHLRPQKHNSTYPOQPOPYAQRGPRPYTYTTPPRMRQRSAA
DLKSGSAV"

Query Match	15.48;	Score	361.8;	DB	10;	Length	4117;
Best Local Similarity	58.58;	Pred. No.	1.1e-91;				
Matches	891;	Conservative	0;	Mismatches	687;	Indels	57;
						Gaps	6;

OY	219	gaatctcccaactgtcgagacaggaagactatccaagatctcgagtcggtct	278
Db	442	GGAGGACCCACGTTTGGATCTCTAACACACAGGTGTATGAGCAATAGAGAGGAAATGCT	501
OY	279	gataatgcatccaagggtlataacaattgtactctcgactcaggtccagccggttcgg	338
Db	502	GCTGATCATCCTTTGAAGGCTACCAAGTGTGCAATCTTTTGCTTACGGTTCAGAGGGGGGCTGG	561
OY	339	caagcccatccaatgaatgagg-----ctatggcaagaagcatgtgctgtaccggagat	392
Db	562	CAGGTCCTACACCAATGATGTGGGGCGGCAGGAACCGGGGAGCAGGGCAATTGATACCTCACT	621
OY	393	ttgcgaagacatbgttcgcygylatthaatgaaatcgcaagaagacaagaacccaattgcac	452
Db	622	CTGCAGGAGACCTCTCTCTCGAGTTAATG---TGAAACGAGTGTCTCAACTTTCTACTC	678
OY	453	cgtcgaatttcgtactctggaaatttaacatgaagaagatgtcgagatctgctgtaatcgtc	512
Db	679	TGTGAGGTGAGACTATATGAGATATTATTCGGAACGAGATACGAGACTCTTTGAACCCCAA	738
OY	513	gacaaaagggaatctcaagatctcgaaacacccgctcgagccgcccactcagtcggaagactt	572
Db	739	GAGTGGGGGCTCTGTGGCGGGGAGCACCCCACTCTGGGCCCATATGTATCAAGACTT	798
OY	573	ggcgaaagctgtctgtcgatcatctccaagaatctgaaatctcatgtatgtgaggaacaa	632
Db	799	GTCATAGTTGGCTGTGACTTCTCTATGCAGAACTTGCTGACTCAAGGACTGTGGAAATTA	858
OY	633	agccagaacggtltccgcgcaacaacatgaagacatcagatcgatcccaagccggtctt	692
Db	859	GGCAAGAACTGTGGGTGCGCACCAACATGAAGAAACAGCAGGCCCTCTCCACGCTGTCTT	918
OY	693	caacttgaaccttgacgcgcaaaagtggcatgatgaagagaccaaattgacacagagaagt	752
Db	919	TACATATCCTCTTTACCCACGCGCTCCCAAGACAGCTTACTGTGACTGGATTCAGAAAAGGT	978
OY	753	tgcgaagatcagctcgtgtagatttgccggtgtcttgagcgagacaagctccacagagctac	812
Db	979	CAGTAAGATCAAGCTTTGTGTAACCTTGCCGGAGCAGCGGGCCGACTCTTATAGGGGCTCG	1038
OY	813	tggagcgacttgaggaagggtcgagagatcaaacgcgtcactcttcgaacctaggtcgtgt	872

Db	1039	GGGCATGGCTGGAAGGAAGGTGGAACATCATTAAGTCCCTGACTACTCTAGGGAAGCT	1039
QY	873	gatttcagcgcgtacgcgatatagtcgtcgtggaaaaacagaagaatcagltagacctta	932
Db	1099	GATCTCACCCCTGGAGAGATTGGCAATC---AAAGAGCGGAGAGTGGCACTTCATCCCTTA	1155
QY	933	ccgagattccgttactgacgtgtgtctctcgaagagccctcttggagggaaactgtatgcgc	992
Db	1156	CAGAACTCTGTGCTTACCTTGGCTACTCAAGGAAATTTGGTGGGAATCTCAACGCACAGC	1215
QY	993	catgtatgcgcgaatllcgcctctgtatataactlttgaagaagactctagtaaccttcg	1052
Db	1216	AATGATTCACACCCCTGAGTCCCGCTGACATCAATTCAGGAGAGACACTCAGACTCTCAG	1275
QY	1053	ataitgcggaactctgcgaaggaatcaagaaccgcgaagtgtgtcatatgaagaccgaagc	1112
Db	1276	GTACCCAAACCCGACCAAAACAGATCCGATGCATATGCTATCATCAATAGAGACCCCAATGC	1335
QY	1113	gcggaatgccgcgagtttgaagaggaaggaactcgcgcagcttgagagaaact-----	1163
Db	1336	CGGCTTATCCGGGAGGCTGCAGAGGAGGTGGCCGGTTGCGGGAACTGCTATATGGCTCA	1395
QY	1164	-----ccagagcagttgtgtgaaggttgaaggtgtgtgcagggaggttcttgcgcgca	1212
Db	1396	GGGGCTTTCACCCCTCTCTAGAGGCTCTAAAGGTGGAAAGGGGAGTCCGAGGTGT	1455
QY	1213	gttgcggaatctgtataccgcgcgaagccgcctctcagagaagaatctgttgcataagag	1272
Db	1456	TCTGCTGCTGCTGATCATCTCCCTCCCTGCCCCAGCTTAACTTCCCTCCACATTA	1515
QY	1273	ccggaatgcgacagtcagaagaatctgagcaag-----gcagaat	1310
Db	1516	CGCGCAGCTGGAGGCCGTCAATTTTCACCAATGCTGAGCCCCAGATTTGGGCGTGAAGAGC	1575
QY	1311	cgtaggaagaacttgacaaggttgagaagctctatcgtggatctcaatcagacctgggaaga	1370
Db	1576	CATGAGAGGCTGTGAGGAACCAAGAAATTAATAGCTGAGCTGAATGAGACTTGGGAGGA	1635
QY	1371	gaagcttgcgaagccgcgaggaagaatccacaaggaagaagaagcgcctcgcgagagcttgcg	1430
Db	1636	GAAGCTCGGTAAAGACAGAACTCTGAGGATGGAGAGAGAGAGATTTGCTGGCTGAAATGGG	1695
QY	1431	tatcagcatcgtaaaaagggcttctgtgtgccttaaccactccaagaatctgacacatctagt	1490
Db	1696	CTCTCCGGGAGATGGGCTACTGTGGGTGTCTCTCCAAAAAAGACTCCCACTGGT	1755
QY	1491	caacttgcgcgatcatcctctctcttgcgtctgagtgctgtgtctacaacatcaagcccgga	1550
Db	1756	AAATCTGAATGAGGAGCCCTTGATGTCTGAATGCCCTTCTTCAACATCAAAAGATGGCT	1815
QY	1551	gacaaggtttggaacgttcaaccaagaatacaaaagcgaagaattctgtcaagcgttgcga	1610
Db	1816	CACCAAGGATTGGCCAGGTAGATGTGACATCAAAAGTTGACTGTGCGCCAGTTCAATCCGGAGCA	1875
QY	1611	gatctgaagaagacacgtgtcctgtttgaaa---atgtgacaacgltgtgacacatcgtgcc	1667
Db	1876	GCACTGCCCTCTTCGCAAGCATCCCTTAGGCTGATGAGAAAGTGAATGATGCACATTGGAAAC	1935
QY	1668	aaaagaaagagcttgcgttcaatgttgaagcgcgttgcgaatcgcgaagcctactgcctcgcg	1727
Db	1936	TTGTGAAGGAGGCTGAGACATATCTCAATGGAAACTTGTGACGGAGGCCGTGTGTCTGAA	1995
QY	1728	caagcgctacaagatcatcctctgggcgatttccacatlltgcattcaaccatcgcgaaga	1787
Db	1996	GTCAAGGCAACAGGATGTATGTATGGGCAAGAAACCATGTGTTCGCTTCATCAACCCGGAGCA	2055
QY	1788	agctcgtgcggaagc 1802	
Db	2056	AGCAAGGCTGGAACG 2070	

RESULT 11
AB074017

LOCUS AB074017 3303 bp mRNA linear ROD 16-JAN-2002
DEFINITION Mus musculus klf1c mRNA for kinesin superfamily protein 1C,
complete cds.
ACCESSION AB074017
VERSION AB074017.1 GI:18181920
KEYWORDS
ORIGIN
SOURCE Mus musculus
MUS musculus
Mus musculus cDNA to mRNA.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (sites)
AUTHORS Nakajima,K., Takei,Y., Tanaka,Y., Nakagawa,T., Nakata,T., Noda,Y.,
Seyou,M. and Hirokawa,N.
TITLE Molecular Motor KIF1C Is Not Essential for Mouse Survival and
Motor-Dependent Retrograde Golgi Apparatus-to-Endoplasmic Reticulum
Transport
JOURNAL Mol. Cell. Biol. 22 (3), 866-873 (2002)
MEDLINE 21644475
REFERENCE 2 (bases 1 to 3303)
AUTHORS Nakajima,K. and Hirokawa,N.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2001) Kazuo Nakajima, University of Tokyo,
Department of Cell Biology and Anatomy, 7-3-1 Hongo, Bunkyo-ku,
Tokyo 113-0033, Japan (E-mail:kazuom.u-tokyo.ac.jp,
Tel:81-3-5841-3336, Fax:81-3-5689-4856)
LOCATION/Qualifiers
FEATURES
source 1..3303
/organism="Mus musculus"
/db_xref="taxon:10090"
1..3303
/gene="Klf1c"
1..3303
/gene="Klf1c"
/codon_start=1
/product="Kinesin superfamily protein 1C"
/protein_id="P88386.1"
/db_xref="GI:18181921"
/translation="MAGASVAVAVRPFNARETSODAKCVSNQMTTSLINPKQSK
DAPKSFEDYSYMSHTSEDEPOFASOOQVRRDIEEMILHFEYANCIFAYGTGAG
KSYTMGRQEPGQGIPOLEDSEFNSHVNQSOISYSAFVMEATCEERDLNP
KSRSLVREHPILGYPVODLSKLAIVSYADIALMCGNARFVATNNMETSRSR
AVPTVTFORSHOILTGLDSBKYSILVNLAGEBNDSCARMLKEGNINIKRSR
TLGKIVISALADLOSRRKRSDFPIPRDSVLYTLKLEENSGNSRTMIALASADINTEE
TLSTLRADRTKQIRCNAINEDPNARLIRLOEVARLRLDLAOGISASALGLAV
EEGSPGVLPASSPPAPASPSPPHNGELSEPSAEPOIGPEEMERLQTEKI
IABLNEYEERLKEKTEALRMEERALLAEMGVAREDEGTVPFSPKTPHLYNEDP
LMSCLLYHIKDGTVRYGOVDYDIKLGOFIREOHCLEFRSIPQDGEVMTVLECEGA
ETVYNGKLIMELVLYKSGNRYVMGKNHFRNHEOARLREBERGVPPRPGSPSPVW
NPAQKELLEOQIDIKLEMERLQDLLENQTRKEKEEDLLLEQRLADSQEDSDK
RSCESWRLLISLKEQLPPTTVONIYKRGLPSSGKRAPRVQIIPQRRLOQKDPK
WATWADLKQAVKEICYEVALADRHRAEIEGALAKMRLLCTYGEPEPGAMRA
VARVMDTVEEEGGGGGSEEGARGAEVEDRAHIDKLTGLIQEKKLONSQKRE
LOALRDLRMERYIPLTODEDNDESGALVTWAPRPGPAVEETVDPDHSPAYRPS
PLLSMERVSRPQRESEMPFGRRLRMKQORLROGLGAGRGGGLRRPARVPFH
DCKLRPFKSPQRESEMPFGRRLRMKQORLROGLGAGRGGGLRRPARVPFH
SLDGSRSRGSGSTQPEPQRLRPQKHNGYPOQPPSPAQRPGPRYPPTTYPRRMROR
SAPDLKSSGAV"

BASE COUNT 757 a 930 c 1013 g 603 t
ORIGIN

Query Match 15.4%; Score 361.6; DB 10; Length 3303;
Best Local Similarity 54.5%; Pred. No. 1.2e-91;
Matches 936; Conservative 0; Mismatches 714; Indels 66; Gaps 8;

OY 151 agtggcaaaactatattgatgagcccgaaagcattgcttcgacggtcgtatggtc- 209
DB 115 AATCCCAACAGAGCAAGATGCCCCAAAGCTTCACCTTGACTATTTTACGCTCG 174
OY 210 -ccttgcaagaatgctcccaactatcgagacaggaagaccttccaagatctcga 267
DB 175 CATACTTGCGTGGAGATCCGACGTTTGCATCTCAGCAGCAGCTTACCGACATAGCA 234

OY 268 gtcccgcttcgtaaatgcatcacaagggtataaacaattgatacttcgacctacggtcag 327
DB 235 GAAAGATGCTACCTCCATGCTCCCTTGAAGGCTACAAATGTGTGATCTTTCGACGGGAG 294
OY 328 accggttcgggaagtctcattcaatgatggg-----ctatggcaaggagatgggtg 381
DB 295 ACGGGGGCTGGGAATCTCTACACCATGATGAGGGCGGACAGACCGGGGACGAGGCAAT 354
OY 382 atcccgagatttcgcaagacatgtcccggtatcatcgaatgaacatgcagaagacaagac 441
DB 355 GTGCTCAGCTGTGAGAGACCTCTTCTCTGAGTCATGTGAACCATGACAGAGTCCAG--- 411
OY 442 ctcaattgcaccgctgaaagtttcgtaacttggaaattcaacaatgaaacgagtgcaagattg 501
DB 412 CTTCCTATTCAGTAGAGGTAGCTATATGAAATCTATTGGACGAGTACGAGTACCTTC 471
OY 502 ctgaatcgtgcacaagaagggaatcgaagtcggaagacacccgttcgacggccctac 561
DB 472 TTGAACCCCAAGATGCGGGGCTCTCTGCGGGTCCGGAGCACCCCATCTAGGCCCTGAC 531
OY 562 gtggaagacttgcaagcgtgctcgatcatcacaagaatcgaatcgaatcgaatcgaatcga 621
DB 532 GTGCAAGACCTGTCTAACTGCTGCTGACTTCTTATGCAAGACTTGCAGCTTATGAGAC 591
OY 622 gagggcaacaagaacgaagacggttcgcccacaacaatgaaacgagacatcagatcgatcc 681
DB 592 TGTGGAAATTAAGCAAGAACCCTGGCTGCCACCAACATGAAGACAGACAGCGCGCTCC 651
OY 682 caagcgcttcctacattgacattgcagcaaaagtgcataatgaaagagaacaaatgagac 741
DB 652 CATGCCGCTTATCTATGTCTTTACCCAGCATGCCATGCCATGACCTTACTGCGCTGGAT 711
OY 742 acaagaaggttcgaagaatcagatcgtgataattgacggttcgagcgaaacagctcc 801
DB 712 TCAGAAAGAGTCAATGAATGATGACTTGTGTGACACTTCGCCGGAAGTGAAGGGCTGACTTC 771
OY 802 accggaagctactgagcgcgactgaagagaggtgcagagatcaaccgctcattcgacc 861
DB 772 TCAGGGGGCCGGGGGATGCTCTAAAGGAAGCGCCCAACATCAATTAAGCTCTGACTGAT 831
OY 862 ctgaagctgtagatgacagcgctagcgaatgtagtctgtaggaagaagaagaatcag 921
DB 832 CTGGGGAAGGTATCTCAGCCTCTGCAGATTTGCAATC---AAAGAAAGGGGAAGTCGGAT 888
OY 922 ttagtaacttcaaggaattcgttactgaacggttccttcgaagactccttcggaagcaca 981
DB 889 TTTATCCCTTAAGAGACTCACTGCTTACTGCTGCTGAAGAGATTTGGGTGGGAAC 948
OY 982 tcgataccgcgcaatgctgcgcacattcgcctcgtgatataacttgaagaagactctc 1041
DB 949 TCTCGACAGCAATGATTGCAAGCCTGAGTCTGCTGACATCAATTAAGAGAGACACTC 1008
OY 1042 agtacccttcgtagtgcgactctgcgaaagcaatcaagaacaaacagcgatgtagtaagaa 1101
DB 1009 AACACTCTCAGGTATGACAGACCGCACCAAGCAGATCCGATGCAATGCTGTCACTCAACGAG 1068
OY 1102 gaaccgaacgcgagatgaccccgagttgaaaggaaggaactcgcgacgtgcgaggaacaa 1161
DB 1069 GACCCCAATGCCGCGCTTATCCGGGAGCTGCAGAGAGAGAGTGGCCGCGTGGGAGACTCG 1128
OY 1162 ctccagagcagtggtggaagtgagtgagtgcaagaagttctcgcgggcagtgag--- 1218
DB 1129 CTGATGGCTCAGAGGCTGTCAAGCTCTCTCTYAGGAGGTCTTAAGGTGGAAGGGGAGT 1188
OY 1219 -----gaatcgataccgcccgaacacgcccgttcgaagaagcaa 1254
DB 1189 CCTGAGGTGTTCTGCCCCCTGCATCATCTCCCGGCCAGCTTCACCTCATCTCTCT 1248
OY 1255 atcgtgctgattcaagcagcgagatgcagacatcaagaanaatgagcaagaacatcg-- 1312
DB 1249 CCCCACATTAATGGGAGAGTGAAGCCCTCATCTTCACCCAGTGCAGGCCACGATTTGGG 1308
OY 1313 -----tgagcaactgaacagagtgagaagctctatcggatctcgaatcag 1359

Db 1309 CTTGAGGAGGCCATGAGAGGCTGACGAGAAACAGAGAAATCATAGCTGAGCTGAATGAG 1368
1360 acctgggaagaagacgtgcgcgaacacccgaggaatccacaaggaagacgagcgctc 1419
Db 1369 ACTTGGGAGGAGGAAGCTTCGGAAGACAGAGGCTCTGAGAAAGGAGAGAAAGCATTTCTG 1428
1420 gaagagctggg-----tatcagcatcgaaaaaggcttctgtgccccttaccactccaaa 1473
Db 1429 GCTGAGATGGGCGTGGCTGTCCGGAGGATGGCGTACTGTGGCTGTCTCTCTCCAAA 1488
1474 gaaatgcacatctagtaactctgagcgtatgctctcttgcgtgagtgctctgtctac 1533
1489 AAGACTCCCACTGTAATCTGAATGAGAGACCCCTTGATGTCTGAATGTCTCTCTAC 1548
1534 aactcaagcccgagcagcaagaaggttggaaagctcaaccaagatacaagaagcgaatt 1553
Db 1549 CACATCAAGAGTGGCGTCCAGAGGTTGGCCAACTAGATGTGACATCACTCACTGGC 1608
1594 cgtctgaacggttcgaagatcctgaaagacactgtacgtttgaaatgtgacaacgtt 1653
Db 1609 CAGTTCACTCCGGAGCAGCATTTGCTCTCCGACATCCCTCAGCTGATGAGAGAGT 1668
1654 gtg--accatcgtgccaaacgagagcgtgctgctatgctgaaacgagcgtgacac 1710
Db 1669 ATGTTACTCTGAGACCTTGTGAGAGAGCTGACATATGTCAATGAGAACTCTGTGACG 1728
1711 aagctactcgcctcgcagcagcgtgctacagatccctcgcgcgtttccacatttctga 1770
Db 1729 GAGTGTCTGCTGTAAGTCAAGAAACGAGATGTGAATGGGCAAGAACATGTTTCCGC 1788
1771 ttcaaccatccggaagaagcgtcgtgcgaacgca 1806
Db 1789 TTCAATCACCCGGAGCAGGCGGCTGTGAACGAGAA 1824

RESULT 12
CEL0104A
LOCUS 5170 bp mRNA linear INV 02-DEC-1993
DEFINITION Caenorhabditis elegans kinesin-related protein (unc-104) mRNA,
complete cds.
ACCESSION M58582
VERSION M58582.1 GI:156477
KEYWORDS kinesin heavy chain; kinesin-related protein.
SOURCE Caenorhabditis elegans (strain N2) cDNA to mRNA.
ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 5170)
Otsuka,A.,J., Jeyaparakash,A., Garcia-Anoveros,J., Tang,L.Z.,
Flisk,G., Hartshorne,T., Franco,R. and Born,T.
The C. elegans unc-104 gene encodes a putative kinesin heavy
chain-like protein
JOURNAL Neuron 6 (1), 113-122 (1991)
MEDLINE 91097805
FEATURES
source
1..5170 Location/Qualifiers
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
1..5170
/partial
/gene="unc-104"
1..5170
/gene="unc-104"
111..4865
/gene="unc-104"
/standard_name="uncoordinated-104"
/note="putative; putative"
/codon_start=1
/product="kinesin-related protein"
/protein_id="AA03517.1"
/db_xref="GI:156478"
/translation="MSSVKVAVRPRPNQREISNTSKCVLQVNGNTTNGHSINKEN

FSPFDSHYSWFARNDPHEITQKQVYEELGEMLEHAFEGYNVICRPAFGQTSKSGSYT
MMGKANDPDEMGIIPRLCNDFARIDNNNDKQVQVSEVSYMEIYCEFYVDLNPNSG
GNLRVREHPILGTYVDLITKMAVCSYHIDCNLDENKATVATMANSTSSSHAVF
TVITQKRCADSNIDTERKSTISYDIAVSEALNENAGNSKTMLAALSPADINPDET
LVTSKLAESTKRRKSKSVYIPYRDSYLVTLREANIGNSKTMLAALSPADINPDET
LSTLRADRARQIVCOAVVEDNPANLIRLENEVYIKLRILLDKIDIVDVOETGK
HKRGPKLPAPHEOLEKLOESEKIMEIGEITWQKLIHEEIKROEHEIRDLGLCA
EDGTTLGVFSPKLPILVNLNEPDLSECTIVYLKGVSVGRPEAREHREDLGSSEA
ILHCEEFINEDNVTLMKPNASQYINGVOYQKIDKMEKMLEMSQVREVELEO
EARSRHNLAAITAPQIDPMKAOELLQIDKIDKMEKMLEMSQVREVELEO
KMTQUTREYSMLBNLOKOVDAOSTSGGSGTWEGBRMLTSLLEPPELKMYSOK
RVVLKAIAIKRWYHFTSVRODLWGNMIPVEANAIVSEVKKQVQFALFTDPMYSP
PPDLPEDEDLTRPYKTVVAIOVDLKNGAHYMSIEKLRQLARMRYPTJDEM
SPADGPMDALAGTDFYDRPFMRVAGAFYLLNNLINVLIKRVAVNPEGEK
GYLKVAIPEVQKDEVINOKKGVROTAKLHRKEDPLKSHKNGEPLSAPPEHJOE
EVEFCRRVYVLOAIDVADYISDVFCOENPFLRDEAPRPEPMNSKSPLEFETONH
IKMSKTFLLYHHPPIIFEVFGHFPQPSDEPNFERNONSALGRSLTRLPQPSLVIS
TPVSKSKANAPIGNNNASVSKHDLWFEICELANNEVVPYIVHAKLQPSLVIS
LHOGIORIKITITCHEKELKMKDCQELVGRIRAGEMAGSDVDVLSIGLPGETM
EFSMDORTFPOFPAAMDSILHNSPLINRVENYDQIYMTLSAVNELDGCQAPVTFD
QELSDSEGSNITSVPSDKSLISRSRSDILCKQSKSDNLASNDIVNLDGMSK
LQSLRIQNLINLIVPEVLEERVGVVSKRGFMNLEKTOGMYTRMYIVRBPYILFED
DRDLVIGIILINLARIHSEDDQAMKVYVNTFSVCYNGRFTLMQMPGDEMYDWLYA
INPLMAGOMLHGNOCTLTSPSSSIAS"

BASE COUNT 1752 a 918 c 1064 g 1436 t
ORIGIN

Query Match 15.0%; Score 352; DB 3; Length 5170;
Best Local Similarity 54.8%; Pred. No. 7; le-89;
Matches 918; Conservative 0; Mismatches 680; Indels 78; Gaps 8;

QY 133 gaagagaagcgcgtaaagtggcgaacacattatgatgtgcccgaagcattgctc 192
Db 198 GGAATATGAGACGACATTAATGTCATTCAATTAACAAGGAACTTCAATTT 257
QY 193 gatcgtcgtatgtgctcttgacaagaagtcctccacactgacgagcagagacta 252
Db 258 GATCATTAATTTGGTCTTTGCGAGGAAATGACCCACATTTTATNACCCCAAAACAAATA 317
QY 253 ttcaagaatctcgcagagtcgcgtctctgataatgcatcacaaggttatacaattgtatc 312
Db 318 TATGAAAGCTCGGAGTTGAATGTTGGAACACGATTTGAAGGGTATTAATGTCTGCAT 377
QY 313 ttgcctcaggttcgaccgcgttgcggcgaagtcctatltatgataaggtatgacgaagg 372
Db 378 TTTCATACGGGTCAAAACGATCAGGAAATCATATACAAATGATGGGAAAGCCCATGAT 437
QY 373 c-----atgagctatccgcgcgagatttcgcagacatgttccgcgcgtatga 423
Db 438 CCAAGATGAATGGGTATATATCCACGCTTGTGCAATGATTTATTTGCAAGATGATAA- 496
QY 424 ctgcagaaggaacaagaacctcactgtcaacgctgcgaagttcgtacttggaaattacaat 483
Db 497 --TAACAATGATTAAGAATGTTCAATATTTCTAGAGGATGCTATATGCAATTTATTT 554
QY 484 gaacgaatgagagacttgaatccgctcgcgaagaagggaatccgaagtcgagaaacac 543
Db 555 GACGAGTAAAGATCTTCTGATATCTTAACCTGGAAGTTAACTTAAGGTTTCGTAACAT 614
QY 544 ccgtcgaaccccccactgagtgagagacttgcgaagcgtgcgtcgtgcatcattccaaga 603
Db 615 CTTTACTTGGACCTTACGTGATGACCTTACCAAAATGCGAGTTTGTCTTACACAGAC 674
QY 604 atcgaagaatctcagatgagagggcaacaagcgaagcgttgcgcgcgaacaacatgac 663
Db 675 ATTTCGAATTCGATGAGGAAGAAATTAACGGAATCTTGTGCAACAAATATGTAAT 734
QY 664 gagacatcagtcgactccacagccgcttcaacttgaacttgcagcgaagaatgcatgat 723

Db	735	TCGCATCATCATCAAGATCCCAATCGCGGTATTCACAACTACTACTACTCAAAAAAGCAATTGT	794
Qy	724	gaagaagaccaaaatgagacacagagaaggttcggaagatcagctcgtgtagattcggcgggt	783
Db	795	GCTGATTTCTAAATTTTGGATTACTGACAAACCAATTCAAAATAATTTCTTTGGTGTGATTGGCCGGA	854
Qy	784	tctgaagcgaagaacgctccacccggagctactcggagcggacttgaaaggaggtgcagaagatc	843
Db	855	TCCGAAGAAGCCATTCTTACAGGAGCGCGAAGGTCAACGACTAAAAGCAAGACCAATATTC	914
Qy	844	aaccgctcaactctcgacccttaggtcgtgtgattcgtcagcggctagcggatcgtcgtc--g	900
Db	915	AACAAGAGCTTGACAAAGTGGGGCTTGTAATCAGTAACTTCGAGAAGAATCAACAAG	974
Qy	901	ggaaaacaagaagaagaatcagttcgtaccctaccggaagatccggtactcgaaggtgctctcgt	966
Db	975	AGAAAAAAGTCCAAACAAAGGTGTGATTCTCTATTCGTATATCTTGCTGAGCGTGGCTCTTG	1030
Qy	961	aaggactcctctggaaggcaactcagatbacccgacatgattcgcgcacatctcgcctcgcgat	1020
Db	1035	AGAGAAATATCTGGAGGAATTCGAAAATCGCATGCTCGCGGCAATTATCTCAGCTGAT	1090
Qy	1021	ataactcttgaaggagactctcagtaaccttcgatatcggagctctcggactcggaaagcgataag	1080
Db	1095	ATTAACTTGTGATGAACACTGATGACTTGTGAGATATCCGATAGCGGAACAAATTTGT	1150
Qy	1081	aaccacgagtggtcgaattgaagaacccgaacgcgcggagatgacccggagtttgaaggagaa	1144
Db	1155	TGTCAACGACTGTTTATGTAGAGATCCAAATCCAAACTTATTCGTGAACATCAATGAAGAA	1210
Qy	1141	ctcgcgcagcttgaggagcaaacctccagagcagtggtggaagtggaagtggtgcaggaggt	1200
Db	1215	GTTATCAAACTTCCTCATCTTTTGAAAGATTAAGGG-----	1240
Qy	1201	tctgscgggcagatggaggaaatcgttacccgccgacacgcgcctcggaaagcaaatcgtg	1260
Db	1250	-----AATTGATGTCACAGACGCTACAGGAGACCCCGGAATAACAAAAAAGA	1290
Qy	1261	tcgattcagcagccggatcgtgcacagctcaagaazaaatgacaaagcagaagatcgttgagcaa	1320
Db	1299	CCGAAATTGCCAGC-----TCATGACTATGAGCAAC-----TTGAGAG	1330
Qy	1321	ctgaaaccagagtgagaaagcttalcgagatctcaatacagaccctgggaagaaagctcggc	1380
Db	1338	CTTCAAGATCTGGAATATTCATGCGCTGAATTTGGAAAGACTTGGGCGAGAAATTAATP	1390
Qy	1381	aagaccgaggaatccacaagaagacgaagacggcgcgtcagagagctcgggtc---atca	1434
Db	1398	CATACAGAGAAATTTAGAAAACACAGAGAGAAAGAACTTGCTGATATAGGGCTTTGGCTGT	1450
Qy	1436	gcatcgaaaaggctctcgttgcgtccctaccactccaagaagaa-atgcacacatctagtcac	1490
Db	1458	GCAAAAGTGGAAACAACTTAAGCGGTATTTAGTCCAAAGAAAGTTACCCATTTAGTTAAC	1510
Qy	1495	ctgagcagatgtactcctctctcgtcgtgagttctctcacaacatcaagcccggcgcagaa	1550
Db	1518	TTGATGATGAAGTTCATTTGATGTCAAGTGTCTCATATATATCTTAAAGAAGAGTAACT	1570
Qy	1555	agggtctggaaacgctcaaccacaagatatacaaacgagaaatctcgtctgaacggtctcgaagtc	1610
Db	1578	AGCGTCGGCCGCCCTGGAAGCTGAACATCTGTCCACACATTCCTCTATCTGGGGAAGCGATT	1630
Qy	1615	ctgaaagaacaccttaacgtttgaaanaatgtgacaaacglttfgaacacatcgtgcacaacgag	1660
Db	1638	CTCGAATTAACATTGTGAAATTTATTAACGAAAGATGAAATGTGCAATTTGACATGAACCG	1650
Qy	1675	aagcgtctgtcagtgatgaaacggagtcgaaatcgacaacgactcagctccgcacggc	1730
Db	1698	AATGATCTTGTCTATATTAATAGAAAAACAATGACAACTCTACTGATATTAACACAGGA	1750
Qy	1735	tacagatcatcctcgtggcagattccoaatttctcgatccaacatcccggaagaagc	1790
Db	1758	TCTAAGATTATTTCTCGGTGAACATCACGTTTTCTCGATATATATGATCCACAGGAAGC	1813

RESULT	13																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
--------	----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Query Match	15.0%	Score 352	DB 3	Length 7107
Best Local Similarity	59.2%	Pred. No. 7.5e-89		
Matches 689	Conservative 0	Mismatches 445	Indels 30	Gaps 4
QY 6	gggagctggaataatcaagagctggtgtgtgctggatagcccgcttcaacgcccgaagatcga	65		
DB 301	ggcgaagtcataatgaatcaacaaagtcgcccgtgaagagtgccgaccttcaatgcgcgaataatcga	360		
QY 66	ccgtggcgcaaaatgatttgtcgagatggaaatcaaacatccctccctcc	125		
DB 361	actagatacgaaatgtatctgctgaaatgaaaaacagacagatcgtacgaatctccgc	420		
QY 126	gggtgcccgaagaagaagcgctgaataatgctgcaaaactattatgatatgcccgaagcatt	185		
DB 421	gccactggaaaaaaatagagagaaaaaaca-----ccmaaacatt	459		
QY 186	tgcgttcgatacgtctcgattgtgcttcttgacaagaatctcccaactatggacaga	245		
DB 460	cgcatctgatactacgtcttcttactcatttaaccccgacagacgaacatttgcctccacga	519		
QY 246	agacctatccaagatctcggagatccgcctcttgatataatgatccaaggttataacaa	305		
DB 520	gacatgtgttcgattcgtggcgacgtggaaatcttgatattgattccaggctataatgc	579		
QY 306	ttgtatcttcgctcagctcagaccggttcggggaagctctcttaatgatatggtctatg	365		
DB 580	gtgcattatctgcttccggccacacaggtttcggcaagctctacagatgatggacacca	639		
QY 366	caagaaatgctgctgtgaccccgagatttgcagaacatgtctcggcgattaatgaact	425		
DB 640	ggagagcaaggcgatcattccacgtctgtgtgaccacttcttctggccattagccacaa	699		
QY 426	gcagaaggaacaagaacctcacttgcacctgcgaatttcgtacttggaaatltacaatga	485		
DB 700	atccacaccccgag---cttatgtacaaagtgtaggtgtcctacatgagatctatnacga	756		
QY 486	acgagtgagagacttgctgctgaatcc---gtcgcaaaagggaatctcaaggtccgaga	542		
DB 757	gaaggtccacgactgtgcctgcagatcccaacccgaacaaagtgcttaaggtccggagca	816		
QY 543	cccgctcgacgagcccccactcgtggagagacttggcgaagctgctgctgcatctccaaga	602		
DB 817	caatcttattgggcccattatgttgacggactgtgcgaactgctgtgacattcttatcagga	876		
QY 603	aatcgaaaaatcactatgataatgaaggaacaagaacccgaagcttgcgcgcacaacaatga	662		
DB 877	catcgataaacctcatgacgcgaaggaacaaacgacgaacagaggggcccacacaaatga	936		
QY 663	cgagacatccagctgatacccacgcgcttctcaacttgcacttgaagcgaagaatgcatga	722		
DB 937	tgccgagtcgtgcgcctccacgcgcgtcttccatcatgtgcttcaactcamaatctcacga	996		
QY 723	tgaagagacccaataatgagacacagaagaagtgcgaaagatcagctcgtgatattggcg	782		
DB 997	tcaagcgcagggagtgacgcgcgaagaagtgctccgcattgtcccttggatgttgatggctg	1056		
QY 783	ttcttgaagagcaaatgctccacccgagctacttgaagcgcagctgaaggaaggtgtgcaagat	842		
DB 1057	ctccgagggggctgtgaaaaagcgagctgtttggcgatctgtcicaaggaaggtcttaact	1116		
QY 843	caacgcctcaacttctgaacctagtgcgtgtgtatgtgacgcgtacgagatatgtcgtcg	902		
DB 1117	caacaaatctctgacacacacttggcctgtgctatctcaagttgggcccgaatcaatccaa	1176		
QY 903	aaaacaga---agaagaatcagttagttaacctaccagagttcgtgactgaagtgcttct	959		
DB 1177	caagaagacgggtacgaacaaatttgcttccctacccgcgaattccgtcattccctggctgt	1236		
QY 960	gaagagctccttgtggaggaacatcgatgacgcgcataatgttcgcgcatttgcgtgtga	1019		
DB 1237	gaagaacaaatttgggtggcaattccagaaacttttatgtatggagacaattctcccatcgccg	1296		
QY 1020	tattaacttgaagagactctcgaatccctcgatatatgcgaaactcgtgaaagcgaatcaa	1079		

Db	Accession	Gene	Protein	Length	Score	E-value	Organism
Db	1297	AGATDACTACGAGACGCTTTCCACGCTTGCCTTATTCAGATCGAGCCAAAGCCATTTGCT		1356			
Qy	1080	gaaccacgcagctgctcaatgaagaccgcgaacgcgcgagatgcccgagttgaagagga		1139			
Db	1357	TAACACGCGCTGTGTTCACGAGAAATCCCAATGCCGCGATCACTTGTGTAGCGTGCACACGA		1416			
Qy	1140	actgcgcgacgttgagagacaact		1153			
Db	1417	GGTGGAGACCCCTTAGAAGTATGCT		1440			
RESULT	14						
LOCUS	AB037923		6618 bp	mRNA	linear	ROD 15-DEC-2000	
DEFINITION	Mus musculus mRNA for KIF13A, complete cds.						
ACCESSION	AB037923						
VERSION	AB037923.1		GI:10697237				
KEYWORDS	KIF13A.						
SOURCE	Mus musculus (strain:ICR) 4 week male brain cDNA to mRNA.						
ORGANISM	Mus musculus						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.						
AUTHORS	1 (sites)						
TITLE	Nakagawa,T., Setou,M., Seeg,D., Ogasawara,K., Dohmae,N., Takio,K. and Hirokawa,N.						
JOURNAL	A Novel Motor, KIF13A, Transports Mannose-6-Phosphate Receptor to Plasma Membrane through Direct Interaction with Ap-1 Complex						
PUBMED	Cell 103 (4), 569-581 (2000)						
REFERENCE	11106728						
AUTHORS	2 (bases 1 to 6618)						
TITLE	Nakagawa,T. and Hirokawa,N.						
JOURNAL	Submitted (02-FEB-2000) Terunaga Nakagawa, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113-0033 Japan						
DEFINITION	(E-mail:terunaga@m.u-tokyo.ac.jp, tel:81-3-5841-3336, Fax:81-3-5689-4856)						
FEATURES	Location/Qualifiers						
source	1..6618						
gene	/organism="Mus musculus"						
CDS	/strain="ICR"						
	/db_xref="taxon:10090"						
	/sex="male"						
	/tissue_type="Brain"						
	/dev_stage="4 week"						
	99..5348						
	/gene="Kif13A"						
	99..5348						
	/gene="Kif13A"						
	/codon_start=1						
	/product="KIF13A"						
	/protein_id="BAB16346.1"						
	/db_xref="GI:10697238"						
	/translation="MSDTKAVAVRVRPNRRELTLTKCKVEKMEGQTVLHPPSNTK KQGRKPKVFAFDYCFWMSMDSNTTVAQGVFKLGSLLEKAGQYNACIFAVG QTGSGKFSFMGHAEOGLIPRLCCALFORIALEONSQTFKEVSYMEITNEKVRDL LDPGSGSLREHKVLGYPVDGLSOLAVTSPEDISIMSEGRKSRVATNNESSS SRSAVNIITLTOLYDLOGSGNSKYSVINDLAGSERKSGAAGEBLKESNTN KSUTLGLVSLADQAAGCKKRFVYRDSVLTWLKLDNGSQTSMATISPAD NYEETSLTRYADRAKRIVNAVVDNEPNAVIRELKEVEKLEQSLSKAEAMPPEL NKLEESBKILKELVTWEEKLRTEIAQERQSLSMGSLSTSGIKVGDQKCYLV NKLNDPALNELLYVLIISQGEVAGRLHVEVITGTIPERMAEDSSNSGSELEV NARCVNGTICNTTOLMHGDRILMGNNHPRILPKRRRDLKDEPERESSHEHL DAEASESDPYNTEPAMQEMVYKTLNSNPYCNVYOLEKYLEKRTLEPRLMY ERELBOLROULSPERPSSASDRILATSSOTAOQVMAEBERELRQSLAKREDD QVNTVLEKNNLIDMRDLYQEMKENVBEAKLGRKQDPFEAEENHLIGAVRKGCTQ CLFDVKLOYAVPIIISQGEVAGRLHVEVITGTIPERMAEDSSNSGSELEV DSSEVHVRKCLIKVYIKKATGLPISLNEVRCOYTFMDCSETVAAVYDDVPS POSQDAOYTVYFHSCKDYVYVTEEFLEFLISDGLALFWNGHRAGNGSPWEDSLH VATVTKLRNNNEVTRLEWVLTSLLENLEQYAVLEHQADVDTGVGFQDROGHSRR QVNTVLEKNNELPLNVEALISVSLICVYATRKILQGRDRDYLKDEPERESSDA						

BASE COUNT 1712 a 1708 c 1841 g 1357 t

Query Match	14.7%	Score 346.6	DB 10	Length 6618
Best Local Similarity	59.5%	Prod. No. 2.6e-87		
Matches 690	Conservative	0	Mismatches 439	Indels 30
			Gaps	
QY	16	aatacaagatggtgtgtgtgcgtgtacgcccgtgtcaacgcccggagaatcgaaccgtgtgcga	75	
DB	111	AAGGTAAAGTTGGCGGTCCGGGTCCGGCCCATGAACCCGAAGAGACTGTGAACACC	170	
QY	76	aaatgatctgtgcgaatggaagaaatcaaacatcctcaccctccctccgtgtgtgcga	135	
DB	171	AAGTCCGTGTGTGAATGGAACGAACCAAGACGGTCTTCACCCCTCTCTTCCAAACT	230	
QY	136	gagaaagcgctgaaaagtgycaaactattatgtatgtgcccgaaggaatttgcgttgat	195	
DB	231	AAGCAGGAGAAAGAA-----AACCTCCCAAGGTGTTCCCTTTGAT	272	
QY	196	cgttcgtatctgtcccttgcaca---gaatgtcccaactatgcgagacaagaaccta	252	
DB	273	TATTGCTTTTGGTTCATGGATGAATCGAACAACAAATAATGCGTGTCACAAAGTCGT	332	
QY	253	ttccaagatctcggaggtccgcctcttgtataatgcatccaaggttataacaattgtalc	312	
DB	333	TTCAAGTGCTCTGGGGAGGAATTCCTCGAAGAGCTTTCAGGGATACATCTGTATT	392	
QY	313	ttcgcctcagtcgaacccggttcggcgcaatcctatcaatgtatggctatgtgcgaag	372	
DB	393	TTTGATATGGGCACACAGGTTCTGGAAGTCTTTTCATGATGTGGCCACCGAAGCAG	452	
QY	373	catggtcgatctccgcgggaatttgcagagacatgtccggcgtatcaatgaactgcgaag	432	
DB	453	CTGGGCTTATTCACCAAGGCTGCGTGTCTGTTTCAAGGAT---CGCCCTGGAGCA	509	
QY	433	gacaagaacctactctgcacgcgtgcgaagtttcgtacttggaatttacaatgaagagt	492	
DB	510	AATGATGTACAGACCTTCAAGTCCAGGTAATGTTATGGAATCTTATACGAGAAGTTC	569	
QY	493	cgaagctgtcgtatcc---gtcgaacaagggaaatctcaagtcgcgaacaaccgttcg	549	
DB	570	CGGGACCTTGTGAGACCCCAAGGAGCAGGCGAGTCTGTGAAGTCCGAGAGCATTAAGTTC	629	
QY	550	accggccctcactcgttggagacttgcggaaagtcgtgcgtgcgtatcatctcaagaatcgaa	609	
DB	630	CTGGGCGCGTATGTGATGTGTTGTCTCAGCTGGCTGTACCAAGCTTTCAGGACATTGG	689	
QY	610	aatccatcgtatgagggcaacaagaacgaagaacggttgcgcgcaacaacaatgaacgaca	669	
DB	690	TCACGTGATGTCTGAAGGAATAATGATCTGMAAGGTATGCAACCAACATGATGAGAG	749	
QY	670	ttcagttgatctccacgcgcgtctcaacttgcacttgaacgcaaaagtgcataatgaag	729	
DB	750	AACACCCGCTTCCCAAGCGGTCTTCAACATCAATCAATCACACACACCTGTATACCTACAG	809	
QY	730	accacaatggaacagagaaggttgcgaagatcaatcttgtatatttgcgggttctgtg	789	
DB	810	TTTGGGAACCTCCGGGAGAAAGTCAACGACGATGAGACCTGTGTAGACCTGGCGGGAACCGAG	869	
QY	790	cgagcaacgttccaccggaacttacttgagcgcgacttgaagagggtgtgcagaatcaaccgc	849	

[illegible]

BASE COUNT 1707 a 1265 c 1439 g 1289 t

Query Match	14.0%;	Score 328.2;	DB 9;	Length 5700;
Best Local Similarly	58.6%;	Pred. No. 4.6e-82;		
Matches 677; Conservative	0;	Mismatches 448;	Indels 30;	Gaps 5;

Oy	16	aatacaaggtgtgtgttcggtatgcagccgttcaacgcccggaaatcgacggtgtgcga	75
Dd	13	AAAGTAAAGTGTCCGTCGGGTCCGGCCCATGAAACGACGAACTGGAATCGAATCAAC	72
Oy	76	aaatgatctgtgcggtatggaagaaatcaaacatccctcaacccctccctcgggtgcgaa	135
Dd	73	AAGTCGTCGTGTGAATGGAAGGAATCAAAAGGTCGTGACCCCTTCCTCTTAACACC	132
Oy	156	gagaagggcgctaaagatgagcaaacattatagatgagccggaaggaatttcgttcgat	195
Dd	133	AAACAGGGGAGAAAGG-----AACTCCCAAGGATATTTCCTTTGAT	174
Oy	186	cggtcgatctgttccctttgacaag--aatgtcccaactatgcgagaaaggaacta	252
Dd	175	TATTGCTTTGGTCATGATGAATCTAACATCAAAATATAGCTGTGTAAGAGGTGT	234
Oy	253	ttccaagatctcggagtcctccgcttcgtatgataatgattcaagaagtatacaattgtatc	312
Dd	235	TTTCAAGTCCTTTGGGGAGCAAGAAATCTTGAAAGCCTTTCAGGGGTATATGCTGTATT	294
Oy	313	ttcgcctacgttcagacgcggttcgggcaagtlcctatcaatgatggtcattgcaagag	372
Dd	295	TTTGCATTATGGACAGACAGGTTCGGGAAATCTTTTCATGATGGCCATCTCGACGAG	354
Oy	373	catgctgtatctccgcgagtttggccaagacatgttccgcgfatataatgaactgcgaag	432
Dd	355	CTGGGCTTATTTCACAGGCTCTGCTCTGCTTTATTTAAAGATCTCTT---TGAACGA	411
Oy	433	gacaaagaacctactgtcacgcgttcgaagtttcgtacttggaaattcacatgaacgagt	492
Dd	412	AATGATGCACAGACCTTTAAAGTTGAAGTCTCTATATGAAATTTATATGAGAAAGTT	471
Oy	493	cgaaactgtctgaatc---gttcgaaaaagggaaatcctaaggtccgaaataaccgcgtg	549
Dd	472	CGGGAATCTTTTAACACCCCAAAGGGAATPAACAATCTCTTTAAAGTTGCAAACTTAAAGTT	531
Oy	550	accgagccctacgttggaggaactgtgcgaaagctgtgtcgatcatcttcaagaagaatcgaa	609
Dd	532	TTGGGACCATATGTAAATGTGTTATCTCAACTAGCTGTCACTAGTTTGTAGGATATTGAG	591
Oy	610	aatctcatgatgtagggcaacaagaccagaagctgttcgcgcgcaacaatgaaacgaga	669

Db	592	TCATGATGTCGTGAGGAATAAGTCTCCAGAGGTAGGCTGCTACCAACATGAACGAAGA	651
QY	670	lccagtcgatlcccaagcgcglcttcaacttgcacttgcacgcgcaaaagtgcattgatgaag	729
Db	652	AGCAGCGCGTCCCATCTGCTGTTCACATCATATACACAGACACTTATAGCCCTCGAG	711
QY	730	acccaaatgcaacacagagaaggtctggaagatcagtcctggtagattttgsggtctcgag	789
Db	712	TCCTGGGAAATTCGGGGAGAAAGTCAGTAAAGTCACTTGTAAGACTTGCGGGGTAGCGAA	771
QY	790	cgagcaacgctccacggagacttactgagcgcgacitgaaggaaggtgcagagatcaacgc	849
Db	772	AGATATCTTAAACAGGAGCTGCAGAGAGCCACTGTAAGAAAGCAGCAACATTATACAA	831
QY	850	tcaacttcgacacctgaagtcgtgcatcgattgcagcgctiagcggatagatctgcgggaaacag	909
Db	832	TCGCTTACAACTTGGGGGTGGTATATCATCATCTGGCTGACCAAGCAGCTGGCA--AG	888
QY	910	aagaagaatcagttagtaacttaacgcgagattcggtaactgcgtgacttctgaaggaattcc	969
Db	889	GGTAAAGCAAAATTTGGCTTATCCAGATATTCAGTCTCACTTGGCTGCTTAAAGCAAC	948
QY	970	ttgggaaggaacatcgatgacgcgcgatgtccgcacatttcgcctgcgcgatataactt	1029
Db	949	TTGGGGGGCAACAGCCAAACCTCTATGATAGCCACATAATAGCCGACCCGACAGCAACTAT	1008
QY	1030	gaagaagactcagtlaccccttgatalctgcygactctgcyaaagcgaatcaagaaccacga	1089
Db	1009	GAAAGAACCCCTTCACATTTAAGTATGCAAGACGAGCCAAAGGATTGTGAAACCATCT	1068
QY	1090	gtgtcgaatgaagaaccggaacgcgcggaatgcgagatctgaaggaaggaactgcgag	1149
Db	1069	GTTGTGATGAGGAGCCCAACGCAAAAAGTATCCGAGAACTCGGGAGGAACTCGAGAA	1128
QY	1150	ctgaggaagcaaacctc	1164
Db	1129	CTGAGAGAGCAGCTC	1143

Search completed: July 25, 2002, 05:37:18
Job time: 6591 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 03:48:52 ; Search time 378.07 Seconds

(without alignments)
10681.045 Million cells updates/sec

Title: US-09-235-416-2

Perfect score: 2352

Sequence: 1 atgctggcggcgtgtggaatat.....aaacagcgaagcagaattc 2352

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2352	100.0	2352	20	AAx87656
2	495.8	21.1	5322	23	ABL14183
3	473.2	20.1	8309	22	AA159190
4	382.2	16.2	1416	22	AA160976
5	380.6	16.2	3930	21	AAZ44744
6	380.6	16.2	3930	22	AAAD08139
7	352	15.0	7134	23	ABL12291
8	321	13.6	2275	22	AA531140
9	316.2	13.4	10682	21	AA65550

10	309	13.1	1701	21	AA622377	Human secreted pro
11	279.8	11.9	2248	22	AA503053	Human diagnostic a
12	276.8	11.8	3525	23	ABL09775	Drosophila melanog
13	177.8	7.6	3319	15	AAQ70152	K39 gene of Leishm
14	177.8	7.6	3319	17	AA742166	Leishmania chagasi
15	164.4	7.0	2034	23	ABL16733	Drosophila melanog
16	164.4	7.0	4034	23	ABL16732	Drosophila melanog
17	161	6.8	15532	23	ABL12290	Drosophila melanog
18	134.8	5.7	12075	23	ABL14182	Drosophila melanog
19	124.4	5.3	758	22	AA123900	Human breast cance
20	124	5.3	1407	22	ABA17294	Human nervous syst
21	118.6	5.0	459	21	AA602955	Human secreted pro
22	114.8	4.9	2111	24	AA599916	Polynucleotide enc
23	112	4.8	5884	22	AA545037	cDNA encoding nove
24	110.6	4.7	5368	22	ABA83095	KIAA0591 ovarian t
25	110.6	4.7	5368	22	AA597834	Human neuroblastom
26	110.4	4.7	8848	23	ABL09774	Drosophila melanog
27	107.8	4.6	3572	19	AAV61162	Drosophila kinesin
28	106.4	4.5	1230	24	AA623972	Human HskR15 prot
29	106.4	4.5	4757	24	AA623970	Human kinesin supe
30	106.4	4.5	4790	22	AAH72653	Human cervical can
31	106.2	4.5	1950	22	AAH15830	Human cDNA sequenc
32	106.2	4.5	3720	23	ABL13337	Drosophila melanog
33	104.6	4.4	814	22	AAH05392	Human cDNA clone (
34	104	4.4	9626	20	AA526819	Nucleotide sequenc
35	103	4.4	1053	19	AAV47580	Leishmania antigen
36	103	4.4	1053	24	AA596044	Leishmania antigen
37	102.6	4.4	5607	22	AA545225	cDNA encoding nove
38	100.4	4.3	8083	22	AA570692	DNA encoding novel
39	100.4	4.3	8503	22	AA158253	Human polynucleoti
40	99.6	4.2	520	22	AA181860	Human polynucleoti
41	99	4.2	4550	23	AA584836	DNA encoding novel
42	98.2	4.2	1391	22	AAH89871	Human bone marrow
43	97.2	4.1	4557	21	AA646311	Arabidopsis thalia
44	97.2	4.1	4989	21	AA642969	Arabidopsis thalia
45	95	4.0	1128	24	AA623971	Human HskR15 prot

ALIGNMENTS

RESULT 1

AAx87656 standard: DNA: 2352 BP.

AC AAx87656;

DE 26-OCT-1999 (first entry)

XX

XX Thermomyces lanuginosus kinesin motor protein TL-gamma DNA.

DE

XX TL-gamma: kinesin; motor protein; microtubule; unc-104; infection;

XX neurodegenerative disease; Alzheimer's disease;

KW Parkinson's disease; Huntington's disease;

KW amyotrophic lateral sclerosis; ss.

KW

XX Thermomyces lanuginosus.

OS

XX

XX WO937659-A1.

PN

XX 29-JUL-1999.

PD

XX 22-JAN-1999; 99WO-US01355.

PF

XX 23-JAN-1998; 98US-0072361.

PR

XX (REGC) UNIV CALIFORNIA.

PA

XX Goldstein LSB, Sakowitz R;

XX

XX WPI: 1999-493950/41.

DR

XX P-PSDB: AAY06618.

XX

PT New nucleic acid encoding microtubule motor protein, used for
PT diagnosis of fungal infection and neurodegenerative disease
XX
PS Claim 4: Page 69-70; 75pp: English.
XX

CC This is the DNA sequence of the Thermomyces lanuginosus gene
CC coding for TL-gamma (see AY0618), a novel ATP-dependent, plus
CC end-directed microtubule motor protein that is a member of the
CC unc-104 family and kinesin superfamily. The gene was isolated
CC from genomic DNA by PCR amplification (see also AX87659-65).
CC The invention provides TL-gamma nucleic acids, proteins and
CC antibodies, and methods of screening for TL-gamma modulators
CC potentially useful for treating hyphal fungal infections and
CC involving anterograde axonal transport, such as Alzheimer's,
CC Parkinson's or Huntington's diseases or amyotrophic lateral
CC sclerosis. TL-gamma nucleic acids are also used to identify
CC polymorphic variants, alleles, homologues etc. of TL-gamma, or
CC other motor proteins, to generate protein structural models, for
CC recombinant production of TL-gamma, as antisense molecules, to
CC produce transgenic or knockout animals, and in gene therapy.
CC Detection of TL-gamma allows differentiation between hyphal and
CC non-hyphal fungal infections.

SQ Sequence 2352 BP; 626 A; 564 C; 672 G; 490 T; 0 other;

Query Match 100.0%; Score 2352; DB 20; Length 2352;
Best Local Similarity 100.0%; Pred No. 0;
Matches 2352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtcggcggttgaaatcacaagtggtgtgtcgggttacgcccgttcaacgcccggaa 60
DB 1 atgtcggcggttgaaatcacaagtggtgtgtcgggttacgcccgttcaacgcccggaa 60
QY 61 atggaacgttgccgaataatgtattgttcggaatggaagaatacaacatccctcacc 120
DB 61 atggaacgttgccgaataatgtattgttcggaatggaagaatacaacatccctcacc 120
QY 121 cctcgggttcgcaagaagagcgcgtaaagtggaacaactattatggtgcccgaag 180
DB 121 cctcgggttcgcaagaagagcgcgtaaagtggaacaactattatggtgcccgaag 180
QY 181 gcatctgcttcgcatcggttcgtatgttcctttgacaaagaatgctcccaatgycaga 240
DB 181 gcatctgcttcgcatcggttcgtatgttcctttgacaaagaatgctcccaatgycaga 240
QY 241 caaggaagactattccaagatcgcgaggtccgcttcgtgataatgataccaaggttat 300
DB 241 caaggaagactattccaagatcgcgaggtccgcttcgtgataatgataccaaggttat 300
QY 301 aacaattgtatcttcgctacggttcgaacggttcgggcaagttcctaattcaatgggc 360
DB 301 aacaattgtatcttcgctacggttcgaacggttcgggcaagttcctaattcaatgggc 360
QY 361 tatggcaagaagatgctgtatccgcggaatttgcgaagacatggttcggggtataat 420
DB 361 tatggcaagaagatgctgtatccgcggaatttgcgaagacatggttcggggtataat 420
QY 421 gaacttcgagaagaagaacaccttcgaacttcgaacgttcgaagtttcgttcttggaaattac 480
DB 421 gaacttcgagaagaagaacaccttcgaacttcgaacgttcgaagtttcgttcttggaaattac 480
QY 481 aatgaacgaatgtcgagacttcgtgaatcgttcgaacaagggaatcacaaggttcgagaa 540
DB 481 aatgaacgaatgtcgagacttcgtgaatcgttcgaacaagggaatcacaaggttcgagaa 540
QY 541 caccgctcgacggcccttcgttggagacttggcgaagctgtgctgtgacatcttccaa 600
DB 541 caccgctcgacggcccttcgttggagacttggcgaagctgtgctgtgacatcttccaa 600
QY 601 gaatcgaaatctcatgtagtggggaacaagaagccgaagcgttgcggccacaacaatg 660
DB 601 gaatcgaaatctcatgtagtggggaacaagaagccgaagcgttgcggccacaacaatg 660

DB 601 gaatcgaaatctcatgtagtggggaacaagaagccgaagcgttgcggccacaacaatg 660
QY 661 aacgagacatccagtcgataccacgcccgttctcaactttgaccttgaagcaaaagtggcat 720
DB 661 aacgagacatccagtcgataccacgcccgttctcaactttgaccttgaagcaaaagtggcat 720
QY 721 gatgaagaagccaaaatgagacagagaaggttgcgaagatcagctgtgatatgttcgcg 780
DB 721 gatgaagaagccaaaatgagacagagaaggttgcgaagatcagctgtgatatgttcgcg 780
QY 781 ggttctgagcgagcaagctccacgcgagctactcgagcgcgactggaaggggtgcagag 840
DB 781 ggttctgagcgagcaagctccacgcgagctactcgagcgcgactggaaggggtgcagag 840
QY 841 atcaacgcctcaacttcgaccccttagtgcgtgtatttcagcgcctagggatatgtctcg 900
DB 841 atcaacgcctcaacttcgaccccttagtgcgtgtatttcagcgcctagggatatgtctcg 900
QY 901 ggaaaaacagaagaagaatcagttagtaagtaacccgagatccggttactgacgttgctctg 960
DB 901 ggaaaaacagaagaagaatcagttagtaagtaacccgagatccggttactgacgttgctctg 960
QY 961 aaggaactccttgaggagcaactcgtatgacccgcatgatttcgccaatttcgctgtgat 1020
DB 961 aaggaactccttgaggagcaactcgtatgacccgcatgatttcgccaatttcgctgtgat 1020
QY 1021 attaactttgaagagacttcagtaaccttcgatatcgagacttcggaagcgaaataag 1080
DB 1021 attaactttgaagagacttcagtaaccttcgatatcgagacttcggaagcgaaataag 1080
QY 1081 aaccacgcaatgtgtcaatgaaagacccgcaagcgcggaatgacccggaatttgaaaggagaa 1140
DB 1081 aaccacgcaatgtgtcaatgaaagacccgcaagcgcggaatgacccggaatttgaaaggagaa 1140
QY 1141 ctgcgcgagcttgaggagcaaatcccaatccagaagctgtgaggttggttgcagaaggtc 1200
DB 1141 ctgcgcgagcttgaggagcaaatcccaatccagaagctgtgaggttggttgcagaaggtc 1200
QY 1201 tctgcgcgccagatgtgagagatcgtaccccccacacgcgcgctcgagagcaaatctgtg 1260
DB 1201 tctgcgcgccagatgtgagagatcgtaccccccacacgcgcgctcgagagcaaatctgtg 1260
QY 1261 tcgattcaagcaagccggaatgcgacagtcacaagaaatgagcaagcagaatctgtgagcaaa 1320
DB 1261 tcgattcaagcaagccggaatgcgacagtcacaagaaatgagcaagcagaatctgtgagcaaa 1320
QY 1321 ctggaaccagagtgaggaagctctacggtatcctaataagaccttggaaggaagctggcc 1380
DB 1321 ctggaaccagagtgaggaagctctacggtatcctaataagaccttggaaggaagctggcc 1380
QY 1381 aagccgaggaagaatccacaagaaggaagcggcgttcgagaggtcgggtatcacgac 1440
DB 1381 aagccgaggaagaatccacaagaaggaagcggcgttcgagaggtcgggtatcacgac 1440
QY 1441 gaaaagggtctgttgccttaccactccaagaagaatgccaactatgactgaactgagc 1500
DB 1441 gaaaagggtctgttgccttaccactccaagaagaatgccaactatgactgaactgagc 1500
QY 1501 gatgatccttcttcgtggtgtgtgtgttaaaacataagcccgagcagaagaaggtc 1560
DB 1501 gatgatccttcttcgtggtgtgtgtgttaaaacataagcccgagcagaagaaggtc 1560
QY 1561 ggaacgctcaacccaagatacaagaagcgaatctgcttgagacggttcgaagatcttgaaa 1620
DB 1561 ggaacgctcaacccaagatacaagaagcgaatctgcttgagacggttcgaagatcttgaaa 1620
QY 1621 gaacactgtacgtttgaaaatgttgacaacgcttgagccatctgycacaagagaagct 1680
DB 1621 gaacactgtacgtttgaaaatgttgacaacgcttgagccatctgycacaagagaagct 1680
QY 1681 gctgtcatgttgaaacggtcggtgcaatcgacaagcctactgcctccgacggtctacag 1740
DB 1681 gctgtcatgttgaaacggtcggtgcaatcgacaagcctactgcctccgacggtctacag 1740

OY 1741 atcaccctggcgatctccacatttcgatccaaccatccggaagaagctcgtcgga 1800
|||||
Db 1741 atcaccctggcgatctccacatttcgatccaaccatccggaagaagctcgtcgga 1800
OY 1801 cggcaagaacaactcgtcttcgcacatctcgcacaaagtcagttggttcgctgc 1860
|||||
Db 1801 cggcaagaacaactcgtcttcgcacatctcgcacaaagtcagttggttcgctgc 1860
OY 1861 ccaagcgcctccgacgcagacacttgcaagcggttcggtcggaagcgagctctgc 1920
|||||
Db 1861 ccaagcgcctccgacgcagacacttgcaagcggttcggtcggaagcgagctctgc 1920
OY 1921 tcaagctccttcgacgcagacacttgcaagcggttcggtcggaagcgagctctgc 1980
|||||
Db 1921 tcaagctccttcgacgcagacacttgcaagcggttcggtcggaagcgagctctgc 1980
OY 1961 gaagctctgaagcgagctcctgaaggttgatcaagaagatctcattcgcagatgacgag 2040
|||||
Db 1961 gaagctctgaagcgagctcctgaaggttgatcaagaagatctcattcgcagatgacgag 2040
OY 2041 ttgagtgatcatttgacgagatctcagaagaagcgagctcgtcggtcggttcgga 2100
|||||
Db 2041 ttgagtgatcatttgacgagatctcagaagaagcgagctcgtcggtcggttcgga 2100
OY 2101 gacaaagaaatagcgaatctcgaagatctcgttcgcggtccgtgacaatacattgcacat 2160
|||||
Db 2101 gacaaagaaatagcgaatctcgaagatctcgttcgcggtccgtgacaatacattgcacat 2160
OY 2161 ggaacacattgataatcttcgtcgtcgaaccgcacattactatgcgggtacccctcgtagt 2220
|||||
Db 2161 ggaacacattgataatcttcgtcgtcgaaccgcacattactatgcgggtacccctcgtagt 2220
OY 2221 gatgcgagcgttgacgagctcttttttggtgataagaagtcgaacagatcgctcatt 2280
|||||
Db 2221 gatgcgagcgttgacgagctcttttttggtgataagaagtcgaacagatcgctcatt 2280
OY 2281 gttgattgttgagagttgctgcaacagcaggtcagatgtaagaagccctgtaaaacagcg 2340
|||||
Db 2281 gttgattgttgagagttgctgcaacagcaggtcagatgtaagaagccctgtaaaacagcg 2340
OY 2341 aagcaggaattc 2352
|||||
Db 2341 aagcaggaattc 2352
RESULT 2
ABL14183
ID ABL14183 standard; cDNA: 5322 BP.
XX ABL14183;
AC
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 37031.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
PN
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
PI Venter JC, Adams M, LI PWD, Myers EW;

XX WPI: 2001-656860/75.
DR P-PSDB; ABB70080.
XX
PI New isolated nucleic acid detection reagent for detecting 1000 or more
PI genes from Drosophila and for elucidating cell signalling and cell-cell
PI interactions -
XX
XX
PS Claim 1; SEQ ID NO 37031; 21pp + Sequence Listing; English.
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16175-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
CC
XX
XX Sequence 5322 BP; 1364 A; 1370 C; 1428 G; 1160 T; 0 other;
SO
Query Match 21.1%; Score 495.8; DB 23; Length 5322;
Best Local Similarity 57.4%; Pred. No. 5.7e-138;
Matches 1037; Conservative 0; Mismatches 722; Indels 48; Gaps 6;
OY 20 tcaaggtggtggtgcgggtgagccggttcaagcccgagaatcgacgtgagcgaat 79
|||||
Db 98 ttaagttggcggttgaggttgcgcccccttcaatcgcgcaataagcaggtgcgaat 157
OY 80 gttatgtggtggtgaggaatacaaccatctcaacccctcctcgtggtcggaaga 139
|||||
Db 158 gcatcatcagatgtgcaggaagccacacagccatcaaccatcaaggtgcgccaaca 217
OY 140 aagcgctgaagaatgagcaaatattatgaltggtccggaagcatctgcttcgactcgt 199
218 ca-----agcgtatcgaataagcttcaacttcttact 253
OY 200 cgtattgtcctttgaacaagaatgctccaactatgagcagaagaagaccattccaaag 259
|||||
Db 254 cctactgtgtacatgatacaccagatgccgatttccacacatcgatggtgtcaag 313
OY 260 atctcgagctccgcttctcgataatgatacgaaggtatacaattatcttcgct 319
|||||
Db 314 acattggcgaggaagatgttaccagcactccttcgataagcatgtctgtactcttgcct 373
OY 320 accgtcagaccggttcgggcaagctcctatcaatgattggttcgaagc----- 373
|||||
Db 374 atggtcagactgagagctggaagctgtatcatcatgattggtgagcagcagcgaag 433
OY 374 atggtcgtgataccgagatttgcaggaatgttcggcgatataatgaaactgcgaag 433
|||||
Db 434 agggcataatacctgataatgataagatctgttccatctgcatacaggtatctgaga--- 490
OY 434 acaagaacctacactgacacgtcggaagttcgctactgtaatttaacaaatgaaagctgc 493
|||||
Db 491 ccgataacctcaagctactcgtgtgaggttccatataatgaaatcaattcgagcgagatca 550
OY 494 gagaacttgctgaatcgttcgacaagaagggaatcctaaggtccgagaacacccgttcgacc 553
|||||
Db 551 gggattgtgtgataccgaaanaacaagggaatctcgtgtgagggagacatcccttttg 610
OY 554 gcccttaagtggaagacttggtggaagctgtgtgcgacatctccaagaatacgaataac 613
|||||
Db 611 gtccctatgtggaagattgttccaaactagccgttaccgattaccaggaacattcagacat 670
OY 614 tcatgataagggcaacaagcgaagcgttgcgcgccaacaacatgaaacgaacatcca 673
|||||
Db 671 tcatgataaggaacaagaagctcgaaactgtgcaactactactaacaatgaacgaacacagtt 730

QY	674	gtctgacccaacgcgcgtcttcacatttgcaccttggaacgcgaagaagtgtgcatagtatgaaagagacca	733
Db	731	ccgcgtctctaaagtcgcgtctttaaacaactctctccaccaacgctgcgtacatgtatctatbaca	790
QY	734	aaatgtgacacaagaagaaggtttgcgaagatcagctgcgtgatagattttggcgggtttcttgaaacgag	793
Db	791	atttgaccacacgaagaaggtatccaaagataagaatctgttgatgactttggccgggttcgaaacgag	850
QY	794	caacgtccacacggagactactgagtcgagtcgagacgtgaagagaggtgtccagagatacaacgcgtcac	853
Db	851	cgagatctccactgtgcgcgaagggcaactgcgtctgaaagggagggagcacaatatacaaaactcct	910
QY	854	tttcgaccccttaagtcgttgatcttcgacgcgtctagagtgatgtcgttcggcggaacaag----	909
Db	911	tgcaccacatttggcgaagttatcttcacgcttttggcggagagtggtctccaagaaaaaagaacaa	970
QY	910	--aagaagaatcaagtgtatgataccttaacgagattccgtgtactgtacgctgtgcttgaagagact	967
Db	971	ccaaagaagacgaattttatcttcgcgtacacgtgtatttcgctcttgaaactcgtctgtacgtga	1033
QY	968	cctcttgagagcgaacactcgaatctgcacgcgcacatgttcgcgcactttccgctctgcgtatattaact	1027
Db	1031	actttggagaggaattctcgaaagacgactatgatttggcgtctatctccaacagacagataataact	1099
QY	1028	ttggaagagactctcagtacaccttcgatatactgcggaactctcgcgaagcgaatacaagaacaaag	1087
Db	1091	acgaatgaaacccctccagcacactgcgtactatgcgtatgcgtatcgtgcgaagacaatgttttgaaag	1150
QY	1088	caagtgtgcatatgaagaaccccgaaacgcgcgagatgcacgcgcgggttggaagagagaaactcgcgc	1144
Db	1151	ctgtgtgcatatgaagaatgtgcataatgcgaagtctgattctcgcaaacctcaagagagagagaaactca	1210
QY	1148	agactagagagcaacaactccagagcagaatgtgtgagagttggaagagttgtgtcgaagagagttctgcg	1207
Db	1211	agctctcgagacttactcacaagaacgcgagagcattgaaagtgtcagagaggaacccgacttgcgaag	1270
QY	1208	ggccaggtggaagaaatcgtgaacccgcgcgcgcgcgcgcgtccgagaaagacaactcgtgtcgtatcc	1267
Db	1271	tgtgtgtgtgagagaaagtcgcgagatgcgaataagatgtagctccaccaaaagctcaacgtgtatbaagt	1330
QY	1268	agcagacgcgaatgc--gaaacagctcaagaagaaatgtgcgaagagcaagaaatcgtgtgcgaacattga	1324
Db	1331	cgccacactaaagctcacgtatcgtatcgtatgtgataccacaacgagatcagatgtgtgtgtgcgtacgc	1390
QY	1325	accagagatggaagaagctcctatctcgtgagatctcacaatcaagactcggagaaagaaagctgtgcgaaga	1384
Db	1391	agggccagagcaagaactatctatgtgcagaagcttacaagaagccttggtagagaaactataaagcga	1450
QY	1385	ccgagagaaatcccaaaagaaagaaagacgcgtctgcgagagagctgtgcgtgtatcagatccgaa	1444
Db	1451	ccgagagagattcgcgtgtgcagagagagcgtctcgtccgaaataggggtgtgtgtctgacaag	1510
QY	1445	agg-----gcttgtgtggtcccttaacacttccaaagaagaatgtccaaactagtacaaactgtga	1498
Db	1511	aagatgtgaataacaagctgtgcgtatcttccacacggaagaagactccgcactttgtgcacaactaa	1570
QY	1499	gcgatgtacctctcttcgtgcgtgaagtgtctgtctcaacaactcaagcccgcgcgcgaagaaag	1558
Db	1571	acgagagatcccaactctgtctgtgagtgtcgtcttactacaatcaagaaggggtccaaactcgt	1630
QY	1559	ttggaagaatccaacaagaatacaaaagcggaaatcgtctgtgaacggttctgaaagactctgtga	1618
Db	1631	tgggttaccacatgaaagacaatgtgcgcccaagacatacgtcttcgcgtatccgacactccca	1690
QY	1619	aaggaacattgtgaagtttggaaaatgtggaacaagcttgtgtgcatactgtgcacaacgagaag	1678
Db	1691	aagagcaactgtgaaccttggagaacaagaacagccggtttacacctgtcgtgcacacaagaagtgt	1750
QY	1679	ctgtgtcatatgtgtgaacgcggtgtgcgaatccacaagcactcgcctccgcgcgaacggctata	1738
Db	1751	ctatcatctatgtaaatgtgagcgcgaagtgtgttgaaacggaggttcttaagacgggtcttc	1810
QY	1739	ggatcatcctcggcgacttccacaatttccgatttccgattcaaacatccgcgaagaagactcgtgtgcg	1798

Db	1811	gggtgatacctcgaagaagacacagtgcttcgccttaccacaatccagacgacgcgaat	1870
Oy	1799	aacggaca	1805
Db	1871	tacggaga	1877
<p>RESULT 3</p> <p>AA159190</p> <p>ID AA159190 standard; cDNA; 8309 BP.</p> <p>AA159190;</p> <p>22-OCT-2001 (first entry)</p> <p>Human polynucleotide SEQ ID NO 1393.</p> <p>Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokineic; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.</p> <p>Homo sapiens.</p> <p>WO200153312-A1.</p> <p>26-JUL-2001.</p> <p>26-DEC-2000; 2000WO-US34263.</p> <p>21-JAN-2000; 2000US-0488725.</p> <p>25-APR-2000; 2000US-0552317.</p> <p>09-JUL-2000; 2000US-0598042.</p> <p>19-JUL-2000; 2000US-0620312.</p> <p>03-AUG-2000; 2000US-0653450.</p> <p>14-SEP-2000; 2000US-0662191.</p> <p>19-OCT-2000; 2000US-0693036.</p> <p>29-NOV-2000; 2000US-0727344.</p> <p>(HYSEQ-) HYSEQ INC.</p> <p>Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; Wang Y, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J; Zhao QA, Zhou P, Goodrich R, Drmanac RT;</p> <p>WPI: 2001-442253/47.</p> <p>P-PSDB; AAM40034.</p> <p>Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -</p> <p>Claim 1; SEQ ID NO 1393; 10078pp; English.</p> <p>The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AA138642-AA442213) with nocotropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokineic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, C.N.S disorders.</p> <p>Note: The sequence data for this patent did not form part of the printed specification.</p>			

[illegible]

Accession	Gene	Species	Location/Qualifiers
QY 702	cttgaacgaanaagtgcatatgtatgtgaagagcccaaatatggaacacgaagaagttgacgaagt	761	
Db 841	cttcaaccgaagagcgccatctgacgcagagacccaatataccaacgagaaagttgacanaat	900	
QY 762	caatctgtgtaatttggcgggtcttcgagcgagcaacgltccaccggagctacttggagcgcg	821	
Db 901	caagcttggttgaaactgtgcgtggagcgagcggtctgactcacggagagccaagggcagcg	960	
QY 822	actgaagagaggtgcagaagatcaacgcgtcaacttgcaccctagtcgtgtatggcagc	881	
Db 961	ccctcaagagagagagcccaacatcaacaagtcgtgaccacccttggccaaggtcatctcgc	1020	
QY 882	gctagcggatctgtcgtctggga-----aaacagaagaagaatcagttagt	926	
Db 1021	cttggtctgaatctgactccggagcccaagaagaacaaagaagaagacagatttcat	1080	
QY 927	acctaacccgaattccgtactgacgtgtcctcttgaaagactcctcttggaggaacactgcat	986	
Db 1081	tccgtaaccggaattccgtgtgtgactctgctccctccggaaaccttggcggtaactcaag	1140	
QY 987	gaccgcgcattgtccgcgcatttcgcctgtctgatatattgaagactctcaagtc	1046	
Db 1141	gacagctatgttgcgcagccttgatcctctgacagacatcaactagatgagacacctagcac	1200	
QY 1047	ccttcgatatgcggaactctgcaagcgagatcaagaacacacgacggtgtgtaattgaagacc	1106	
Db 1201	gtctgaggtatgtctgacacggcgccagacgagatccgctgtcaatgtctatcaatgaagagcc	1260	
QY 1107	gaacgcgcgagatgtaccgcgcagatgtgaagaggaagactcgcgcagctgaagagcaactcca	1166	
Db 1261	caacacaagctgtatccgcgcagctggaaggtgtgagtgaccgcgcttgcggactcttgta	1320	
QY 1167	gagcagtgctgtggaagttg 1183		
Db 1321	cgcccaaggtcttggcg 1337		
RESULT 5			
AAZ44744			
ID AAZ44744	standard; cDNA; 3930 BP.		
XX AC AAZ44744;			
XX DT 17-APR-2000	(first entry)		
XX DE Human KLIMP cDNA.			
XX KW KLIMP; kinesin-like motor protein; cytosolic; anticonvulsant; human;			
XX KW anti-Alzheimer; anti-Parkinsonian; antidiabetic; anti-ulcerative; cancer;			
XX KW immunomodulatory; antiinflammatory; anti-AIDS; antirheumatic; treatment;			
XX KW antiarthritic; diagnosis; neurological disorder; vesicular transport; ss			
XX OS Homo sapiens.			
XX Key Location/Qualifiers			
XX FT 78..3389			
XX FT CDS /tag= a			
XX FT /product= "KLIMP"			
XX PN US6013454-A.			
XX PD 11-JAN-2000.			
XX PF 28-SEP-1998;	980S-0162373.		
XX PR 28-SEP-1998;	980S-0162373.		
XX PA (INCY-) INCYTE PHARM INC.			
XX PI Tang YT, Corley NC, Patterson C, Guejler KJ;			
XX DR MPI; 2000-126064/11.			

QY 366 caaggagcgtggtgacccgcggtgattccagagacatgttccggtattaatgaact 425
 Db 665 ggagagcaaggcattccacgtctgtgtgacacactcttctcgccatagccaaca 724
 QY 426 gcagaagacaaagacactcactgtgacgcgtgaagtctgacttgaaattacaatga 485
 Db 725 atccacacccgag---cttatgtacaaaggtggaggtgtcctcatgagatctataacga 781
 QY 486 acgagtcgagactgtgtaatcc--gtcgacaaagggaatctcaagtccgagaca 542
 Db 782 gaaggtccagatctgtctgaccccaagcgaacaaacagtcgcttaaggtgcgcgagca 841
 QY 543 cccgtgcacccgttctgagggacttggcgaagctgtgctgcgcatcattccaaga 602
 Db 842 caatgttatggcccatatgtgacgcgactgtgcagctggtgcacatccatcagga 901
 QY 603 aatcgaaatctcatgatgagggcaacaaagcagaaacaggttgcgcacaaacatgaa 662
 Db 902 catcgataacctcatgacccgaagcaaaagtcacgaacggtggtgcgcacaaatgaa 961
 QY 663 cgagacatccagtcgacccacgcgtcttcacattgaccttgacgcacaaagtggcatga 722
 Db 962 tgcgagtcgtcgtctccacgcgtcttctcagtggtctcactcagatactcaacgga 1021
 QY 723 tgaagagacaaatgacacagagagaaggttgcgaagatcagttctgttagattggcggg 782
 Db 1022 tcaggcagcggagtgagcggcagagaaggtgtcccgcatgtccctggtgattggctgg 1081
 QY 783 ttctgagcagcaacgtccaccggagctactgagcgcgactgaaggaggtgcagagat 842
 Db 1082 ctccgagcgggtgtgaaaacggagcgtgtgtgcatcgtctcgaaggagcgtcctaact 1141
 QY 843 caaccgtcatttcgacctagctgtgtgattgcagcgtcagcgtatgtctgcggg 902
 Db 1142 caacaaatctcgaccacacttggcctgttcacatccaaagttggccgataccaatgg 1201
 QY 903 aaaaacaga---agaagaatcagttagtagtaccttaaccgagatcgtgactggtgtct 959
 Db 1202 caagaagagcgttaacgacaaaattctcctcaccgcatcgtgtcctacgtgctgct 1261
 QY 960 gaagagactcttggagcgaactcgatgacgcgcacatgattgcgcctattcgcctgtga 1019
 Db 1262 gaagaacaaatttggcgaattccagacactgttattgtgtgagcacaatctcgccatcggc 1321
 QY 1020 tattaactttgaagagactctcagatcccttcgatatgcgagactctcgaagcgaatcaa 1079
 Db 1322 agataactcagagagagcgtcttccacgctgctgtatgcagatcgagcgaagcgtattg 1381
 QY 1080 gaaccacgcagtggtcaatgaagacccgaacgcgcgagatgatccgcgagttgaaggagga 1139
 Db 1382 taaccacgctgtgtcaacgaagaatcccaatgcccgcatcattcgtgagctgcgacaga 1441
 QY 1140 actcgcgcagctgagggacaaact 1163
 Db 1442 ggtggagacccttagaagtatgct 1465

RESULT 8
 AAS31140
 ID AAS31140 standard; cDNA; 2275 BP.
 AC AAS31140;
 XX
 DT 04-DEC-2001 (first entry)
 DE Human diagnostic and therapeutic polynucleotide (DTHP) #155.
 XX
 KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
 KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
 KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
 KW respiratory disorder; ss.
 XX

OS Homo sapiens.
 XX WO200162927-A2.
 XX 30-AUG-2001.
 XX 21-FEB-2001; 2001WO-US06059.
 XX 24-FEB-2000; 2000US-0184693.
 XX 24-FEB-2000; 2000US-0184697.
 XX 24-FEB-2000; 2000US-0184698.
 XX 24-FEB-2000; 2000US-0184768.
 XX 24-FEB-2000; 2000US-0184769.
 XX 24-FEB-2000; 2000US-0184770.
 XX 24-FEB-2000; 2000US-0184771.
 XX 24-FEB-2000; 2000US-0184772.
 XX 24-FEB-2000; 2000US-0184773.
 XX 24-FEB-2000; 2000US-0184774.
 XX 24-FEB-2000; 2000US-0184776.
 XX 24-FEB-2000; 2000US-0184777.
 XX 24-FEB-2000; 2000US-0184797.
 XX 24-FEB-2000; 2000US-0184813.
 XX 24-FEB-2000; 2000US-0184837.
 XX 24-FEB-2000; 2000US-0184841.
 XX 24-FEB-2000; 2000US-0185213.
 XX 24-FEB-2000; 2000US-0185216.
 XX 12-MAY-2000; 2000US-0203785.
 XX 15-MAY-2000; 2000US-0204226.
 XX 16-MAY-2000; 2000US-0204525.
 XX 16-MAY-2000; 2000US-0204821.
 XX 16-MAY-2000; 2000US-0204908.
 XX 16-MAY-2000; 2000US-0205232.
 XX 17-MAY-2000; 2000US-0204815.
 XX 17-MAY-2000; 2000US-0204863.
 XX 17-MAY-2000; 2000US-0205221.
 XX 17-MAY-2000; 2000US-0205285.
 XX 17-MAY-2000; 2000US-0205286.
 XX 17-MAY-2000; 2000US-0205287.
 XX 17-MAY-2000; 2000US-0205323.
 XX 17-MAY-2000; 2000US-0205324.
 XX (INCY-) INCYTE GENOMICS INC.
 PA Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
 PI Chen A, D'Sa SA, Amshay S, Dahl CR, Dam TC, Daniels SE;
 PI Dufour GE, Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL;
 PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daifo A;
 PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
 PI Cohen RJ, Hodgson DM, Lincoln SE, Jackson S;
 XX WPI; 2001-502867/55.
 XX P-PSDB; AAU19569.
 XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
 XX enzymes, hormones and receptors, useful in diagnostics and therapeutics
 XX -
 XX Claim 1; Page 369; 522pp; English.

The invention relates to polynucleotides (I) encoding diagnostic and therapeutic (DTHP) polypeptides (II), which include e.g. enzymes, and proteins involved in growth and development and receptors. (I) and (II) may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate DTHP expression. For example, (I) and (II) may be used to treat disorders associated with decreased polypeptide expression by rectifying mutations or deletions in a patient's genome, that affect the activity of the DTHPs, by expressing inactive proteins or supplementing the patient's own production of them. (I) and (II) may be used to treat diseases, for example, cell proliferative disorder, Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma, leukaemia, autoimmune disorders, and respiratory disorders. Additionally, (I) may be used to produce the DTHPs, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) and

CC in disease diagnosis and population studies.

XX
SQ
Sequence 10682 BP; 2817 A; 2443 C; 2602 G; 2812 T; 8 other;
XX

Query Match	13.4%;	Score 316.2;	DB 21;	Length 1082;
Best Local Similarity	59.6%;	Pred. NO. 8.9e-84;		
Matches 616; Conservative	1;	Mismatches 389;	Indels 27;	Gaps 4;

Qy	157	aaaactattatgatggccgcaagcatttgcgttcgactcggctcgatttggctcctttgac	216
Db	307	aagaatccaaaggagctccaaagtccttcaggcttcgactattcctactgtgtccatacc	366
Qy	217	---aagaatgctcccaactatcgagacaggaagaccctattcccaagatctcgagtcctcg	273
Db	367	tcaccggaagatcccgttttgatctcaaacccgtgtgtacaaTgacattTgccaagaa	426
Qy	274	ctcttgataatgcattccaaggttatacaaatTtatcttcgctacoggtcgacacoggt	333
Db	427	atgctcttacacgcctttgaggataataatgctgtatttttgcstatggcgacactggt	486
Qy	334	tcgggcaagtcctattcaatgatgggctatTgcaagg-----agcatggcgatccccg	387
Db	487	gctggaaatctctatacaatTgttgataaacagagaagagcgaggtcgcatcattcca	546
Qy	388	cgaattggcagacacatgttcggcggtattaatgaactgcagaaggacaagaacctcact	447
Db	547	cagttaTgtgaagaactttttgagaaaaTcaatgacaactgtTaatTgaagaaa---	603
Qy	448	TgaccgctgaagtttcgtactctTgaaatttacaatgaacgagTgcgagactTgtcgaat	507
Db	604	tactctgtagagTgagctacatggaaatttactgtgaagagTcacgagattTgtcgaat	663
Qy	508	ccgtcgacaaaggggaaatctcaagTtcogagacaccccgTcgaccggccctcagctggag	567
Db	664	ccaaaacaacgggtaatTtgcgTgcgtgacaccccaactctTggaccctatgtggag	723
Qy	568	gacttgcggaagctgctgcTgcgatcatcccaagaaatcgaaaatctctatgagTgagggc	627
Db	724	gatctgtccaaTgtggcagTtacttctcacagacattgtgacctcaTggaTgctTggg	783
Qy	628	aacaaagcccaagacggtTgcgcgcacaaacaTgaacgagacatccagTcgatccccagcc	687
Db	784	aacaaagccgagacgTggcagctacaaacaTgaatgaacaagTgacgctgccacgct	843
Qy	688	gtcttcaactTgaccttgacgcacaaagTggcagTgatgagagagaccaaataTggacacagag	747
Db	844	gtgttacgattTtttcccccagaagaacacgataTgagaccaaactTtccactTgag	903
Qy	748	aagTgtcggaagatcagctTggtagattTggcggtTctgagcgagcaacgTccaccgga	807
Db	904	aagTgcagTaaatcagctTggTggatctagcaggaagTgaacagctTgattcaactTggt	963
Qy	808	gctactggcgcgactgaagagggTgcagagatTcaaccgctcaactTtcgacccctaggt	867
Db	964	gccaaagggactcgattaaaggaggaacaaTtaataaTgaTctcttacaactTgggc	1023
Qy	868	cgTgtgattgcgcgctagcggga-----tatTgctgctgggaaaaacagaaag	912
Db	1024	aaagTcaatttcagcctTggccgagTgggaactgcaactgacagTgataaaagaagaag	1083
Qy	913	aagaatcagTtagTactctacgagattcggTactcagctggctcttcgaaagactccttg	972
Db	1084	aaaacagattTtatctccTcacgggattcgtTactTactTggctcttcgagaaaaTtta	1143
Qy	973	ggaggcaactcgatgaccgcacatTgattgcgcacatttcgcctgctgataTtaactTtgaa	1032
Db	1144	ggtggcaattctcgactgcaatTggtTgctTgctcTgcggcccgcggaTatacaactcagat	1203
Qy	1033	gagactctagTacccttcogataTgcgggactctgcgagcggaataaagaacacgcagTg	1092
Db	1204	gagactTtggacactctgagataTgcagatcgtgcgaacaaatTaaatgaaTgactTgtt	1263

Qy	1093	gtcaatgaagaccgaacgcgcgcggtgatccgcgcgagttgaaggagggaactcgcgcagctg	1152
Db	1264	atcaatgaggaccaccaactggttcgtgaattaaaggaggaggtgacacgcgctg	1323
Qy	1153	aggagcaaacctc	1165
Db	1324	aaggacctcttc	1336
RESULT 10			
AAE22377			
ID	AAE22377 standard; cDNA; 1701 BP.		
XX			
AC	AAE22377;		
XX			
DT	26-MAR-2001 (first entry)		
XX			
DE	Human secreted protein gene 5 SEQ ID NO:15.		
XX			
KW	Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;		
KW	antirheumatic; antiproliferative; cytostatic; cardant; vasotropic;		
KW	cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide		
KW	fungicide; ophthalmological; gene therapy; autoimmune disease; neoplas		
KW	rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;		
KW	cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;		
KW	angiogenesis; nervous system disorder; Alzheimer's disease; infection;		
KW	ocular disorder; corneal infection; wound healing; skin aging;		
KW	food additive; preservative; ss.		

RESULT 10
AAAF22377
ID AAF22377 standard; cDNA; 1701 bp.
XX
XX
AC AAF22377;
AC
XX
XX
DT 26-MAR-2001 (first entry)
XX
XX
DE Human secreted protein gene 5 SEQ ID NO:15

Human; secreted protein; diagnosis; immunosuppressive; antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective; neoplastic; neuroprotective; antibacterial; virucide; fungicide; ophthalmological; gene therapy; autoimmune disease; neoplasm; rheumatoid arthritis; hyperproliferative disorder; cardiac arrest; cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia; angiogenesis; nervous system disorder; Alzheimer's disease; infection; ocular disorder; corneal infection; wound healing; skin aging; food additive; preservative; ss.

OS Homo sapiens.

XX PN WO200061629-A1.

XX 19-OCT-2000

XX
DE 06-APR-2000-

XX
XX
00-APP-1000.
0000-012960

PR 20-JAN-2000; 2000US-0176931.

PA (HUMA-) HUMAN GENOME SCI INC

XX

XXX

DR P-PSDB; AAB63138.

XX Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
PS Claim 1; Page 419-420; 533pp; English.

PS Claim 1; Page 419-420; 533pp; English.

AAF22373 to AAF22421 encode the human secreted proteins given in AAB63134 to AAB63182. AAB63183 to AAB63231 represent more human secreted proteins and polypeptides homologous to them. Human secreted proteins have activities based on the tissues and cells the genes are expressed in. Examples of activities include: immunosuppressive; antiarthritic; antirheumatic; antiproliferative; cytostatic; cardant; vasotrophic; cerebroprotective; nootropic; neuroprotective; antibacterial; virucide; fungicide; and ophthalmological. The polynucleotides and proteins can be used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for

PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -

PS Claim 1: SEQ ID NO 23807: 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3525 BP; 896 A; 914 C; 1009 G; 706 T; 0 other;

Query Match 11.8%; Score 276.8; DB 23; Length 3525;
Best Local Similarity 57.7%; Pred. No. 3.4e-72;
Matches 576; Conservative 0; Mismatches 407; Indels 15; Gaps 4;

Qy 179 aggcattgcttcgagcgttcgtattgtctctttgacagaagaatgctcccaactatgcga 238
Db 98 acgacttcaccttcgactactccctattgttcattcgatcgagagatcgccacttcgca 157
Qy 239 gacaggagacactattccaagatctcgagtcgcgcgtctcttgataatgattcaagggtt 298
Db 158 cccaggacaggtttacagcagatctgggcaatgatgtgtgattgcgcctacgaagat 217
Qy 299 atacaattgtattcttcctacgagcagcgttcgggcaagtcctattcaatgatgg 358
Db 218 acaatgctgtgttttccctacgacagactggctctggaagacctaccatgatgg 277
Qy 359 gctatggcaaggagatgctatcccgcgatttgcagacatgttcggggtatta 418
Db 278 gcacgcaacaactcccgccctaataccgcatctcgagggagctcttgcgcgcgcgcgc 337
Qy 419 atgaactgcagaagacaaacctcaacttcacgcgtcgaaagtcttcgtacttggaattt 478
Db 338 gtg---tgggccaggagtcgggcaactggtctacaggacgcgatcgccagctatcgagatct 394
Qy 479 acaatgaacgagtcgagacttgctgaatccgctgcagacaaaggggaat---ctcaaggctcc 535
Db 395 acaacgaacgggtcaaggatctctgctggcgacagagcactggcgatggctaaagggtgc 454
Qy 536 gagaacacccctgcacgcccctacgttgaggacttggcgaagctgctgcgatacat 595
Db 455 gggagcatcgcagcgtgggtccctatgtggaaaacctgtcacagcatgctgtctctgatt 514
Qy 596 tccaagaaatcgaaaatctctatgtaggagggcaacaagccagaaagcttgcgcgcacaa 655
Db 515 tcatgagattcaggaaatgattcgaagggaatggaaatgcgaacgcacacgcccagcaca 574
Qy 656 acatgaacagacatccagtcgatcccaacgcgtcttcaactttgaccttgacgcgcaaaagt 715
Db 575 acatgaacagatcacgagcgcgcgcagtcacgcgcacatcttcacgatcacatttgcca---- 629
Qy 716 ggcattgatgaagacacaaatggacacagagaagggttcgaagatcagctgtgtagatt 775
Db 630 -ggctgttcttcagtaatgacatgcacagcagagacagtcgaaagatccatttggctgatt 688
Qy 776 tggcgggttctgagcagcaactccacgcgagctactggagcgcgactgaaggaggggtg 835
Db 689 tggcaggcagtgagcgtgccaatgcaacgggagcaacgggtcagcggctgaaggagggcg 748
Qy 836 cagagatcaacgcgtcaacttcgacccctaggtctgtgattgcagcgcgtacgggatatgt 895
Db 749 ctcatcatacaaaaactcgtcgtgactcgtgggtgagcgtgactcgtcttggcggagcaaa 808
Qy 896 cgtcggggaaaacagaaa---gaagaatcagttagttacattaccgagattcgttactgacgt 952

Db 809 cgggtcggcgttcacaaacagcagcagctcctacattccctatcgactcccaactcacgt 868
Qy 953 ggcttctgaaggactccttgggagcgaactcgatgacccgcatgattgccgcatttcgc 1012
Db 869 ggctgctcaaggacagtttgggtggcaatagcaagaccattatgattccgctcttcgc 928
Qy 1013 ctgctgatatataactttgaagagactcgaactcgaactccctcgatcggaactcgcgaagc 1072
Db 929 cggcgattgcaactacagtgaaacacactcagcagcgtcgggtacgcgaatcgggcgaaga 988
Qy 1073 gaatcaagaacacacagcagtggtcaatgaagaccgcaacgcgcgggatgacccgcgattga 1132
Db 989 acatcatcaacaagcgcgcgttcacacgaggacacaaatgtcaagctgacccgtgagctca 1048
Qy 1133 agggaggaactcgcgcgagctgagggagcaaaactccagagc 1170
Db 1049 gggaggagatcaataaagctgaagctccatgctggcgggc 1086

RESULT 13
AAQ70152
ID AAQ70152 standard; cDNA to mRNA; 3319 BP.

XX AC AAQ70152;
XX XX 08-MAR-1995 (first entry)
XX K39 gene of Leishmania chagasi.
XX Parasite; leishmaniasis; infection; repeat unit; vaccine; antigen;
KW Leishmania chagasi; ss.
XX OS Leishmania chagasi.
XX FH Key Location/Qualifiers
FT CDS 455..3319
FT /*tag= a
FT /product= K39.
XX WO9416331-A.
XX 21-JUL-1994.
XX 10-JAN-1994; 94WO-US00324.
XX 15-JAN-1993; 93US-0006676.
XX (IASY-) IASYS CORP.
XX PA Reed SG;
XX PI WPI; 1994-249402/30.
XX DR P-PSDB; AAR57365.
XX XX Diagnosis of Leishmaniasis - by determining the presence of
XX PT antibodies that bind to a K39 repeat unit antigen
XX PS Disclosure; Page 15-17; 28pp; English.
XX CC The K39 gene encodes a polypeptide which comprises a number of
XX repeated units (described in AAR57366). Detection of antibodies
XX directed against this repeated unit in a patients sample is
XX indicative of leishmaniasis. The antigenic repeat unit can
XX itself be used as a vaccine to protect against infection by a
XX leishmania parasite.
XX SQ Sequence 3319 BP; 669 A; 969 C; 1196 G; 485 T; 0 other;

Query Match 7.6%; Score 177.8; DB 15; Length 3319;
Best Local Similarity 52.6%; Pred. No. 2e-42;
Matches 525; Conservative 0; Mismatches 417; Indels 57; Gaps 4;


```

Db 865 aatgtgatcagtgcttggat-----ggaaaaagcaccacata 906
Qy 928 ccttaccgagattoggtactgacgtggtcttgaaggactccttggaggcgaactcgtg 987
Db 907 ccgtatcggaactccaagttaactcgactctgcaggactccctcgtggttaactccaag 966
Qy 988 accgcatgattgcgcacatttcgctgctgatattaaactttgaagagactctcagtacc 1047
Db 967 accgtcatgtgtgcaccatttagtcacggtatagcaattatatggaaactatatccaca 1026
Qy 1048 ctgcgatgcggactctgcgaagcgaatcaagaaccacgcagtggtcaatgaagaccgcg 1107
Db 1027 cttcgctatgctagtcgtgctaagaataccaataatcgcatgcacatcaatgaggaaccc 1086
Qy 1108 aacgcgcggatgatccgcgagttgaaggaggaaactcgcgcagctgaggagcaaaactccag 1167
Db 1087 aaggatgcttgccttcgtcactttcaggaggagattgccgacttcgcaagcaattggag 1146
```

Search completed: July 25, 2002, 05:44:40
Job time: 6948 sec

.


```
Db 347 cgtgtgcatctttgctctatgggcagaccggggcgtgggaaatcctatatcaccatgatggggcg 406
QY 360 ---ctatggcaagagcatgctgtgatcccgcgatttgcagacacatgttccggcgctat 416
Db 407 acaggagccaggcgagcagggcatcgtgcccgactctgtgaggaccttcttctcgcgt 466
QY 417 taatgaactcagaagagacaaacactcacttgccacgctcgaagtctcgtacttggaaat 476
Db 467 tagtgaacaacagagtctcag---ctatcctactctgtgaggtgagctatattggagat 523
QY 477 ttacaatgaacagagtcgagacttctgaatccgtgcacaaagggaatctcaaggtccg 536
Db 524 ctactgtgagcgggtgacgacctcttgaaccccaagtcggttctctgcgggcccg 583
QY 537 agaacaccctcagcccgccctacgtggaggacttggcgaagctggtcgtgagatatt 596
Db 584 ggagcaccctatcctggcccgctacgtgcagacctgtccaaattgctgtgacctcta 643
QY 597 ccaagaaatcgaaatctcatgtagtgggcaacaaagccagaaacggttgcgcgcacaa 656
Db 644 cgcagacattgctgacctcatgactgtgaaataaagacgagctgtggtgcaccaa 703
QY 657 catgaacgagacatccactgcatccaccgctcttccacttgcacttgacacctgagcgaagt 716
Db 704 catgaatgagaccagcagcgttcccatgcgctctttaccatcgtcttccacagcgtg 763
QY 717 gcatgtagaagaccacaaatgacacagagaaggttgcgaagatcagttcgtgtagatt 776
Db 764 ccatgaccagctcacgggctggaactcggagaaggtcagtaagatcagtttggtagcct 823
QY 777 ggagggttctgagcagacacgtccaccgagctacttgagcgcgactgaaggagggtgc 836
Db 824 tgcctggagtgagcagccgactcctcactcagggcccggtggtggtgaggaaggagc 883
QY 837 agagatacaacgcgtcacttctcccttaggtcgtgtattgcagcgtagcgtatgctc 896
Db 884 caacatacataagctcctgactacactagggaaagtgtctcgtcccttcagatatagca 943
QY 897 gtccggaaaaacagaagaatacagttagttaccttaccgagattcgtactgacgtgct 956
Db 944 atc---aaagaagcgaaagtcgagttttatccccctacagggacctctgtcactgct 1000
QY 957 tcggaagactccttgggaggaactcgtgacccgcacatgattgcccactttcgcctgc 1016
Db 1001 gctcaaggaaatttggggggaactcacgcacagccatgattgcagccctgagccctgc 1060
QY 1017 tgattataactttgaagagactcagtaaccttcgatatgcggaactctgcgaagcgaat 1076
Db 1061 tgacataattacgagagactctcagaccctcaggtatgctgcagccgacccaagcaat 1120
QY 1077 caagaaaccacgcagtggttcaatgaagaccggaacgcgagatgcgcgagttgaagga 1136
Db 1121 ccgctgcaatgccatcatcacgaggaccctaatgcccgtgtattagagagctgcagga 1180
QY 1137 ggaactcgcagctagagagcaaac-----ccagagcagtggt 1176
Db 1181 ggaagtgcgcgcgtcgggaaactcgtggtcaggggactgtcagcctctcgtctgga 1240
QY 1177 ggaaggtgaggtggtcagagaggtct-----ggcgccagtgaggaaatcgtacc 1229
Db 1241 aggcctgaagacggaagaagaggtgtcagaggcgccctccagctgtgtcatctcccc 1300
QY 1230 gccgacacgccctcgtcagagaacaaatcgtgtcgtattcagcagccgagtcgacagtc-- 1287
Db 1301 agctccagtttccacctcatccaccacacacataatggggagctggtgagcgtcatctc 1360
QY 1288 -----aagaaatgagcaagcgagaaatcgtggagcaactggaaccaagatga 1334
Db 1361 ccccaacacggagtcaccagattggcctgaggagccatggagagcgtgcagagagaga 1420
QY 1335 gaagctctatcgggtatcctaatcagacctgggaagagaagctggccaagaccgaggaat 1394
Db 1421 gaagattatagctgagctgaacgagacatgggaggaagactcagcagacagagaagccct 1480
```

```
QY 1395 ccacaagaaacagagaagcgcgctcgaggagctgggttatcagcatc-----gaaaagg 1448
Db 1481 gaggatggagagaagaagcattgctgctgagatgggggtggtccggtccggaggatggggg 1540
QY 1449 ctttgtggcccttacacatccaaaagaatagccacatctagtcacactgagcgtatcc 1508
Db 1541 aactgtggcgctctctctccaaagaagactcccccactggtgaacctgaacgaagaccc 1600
QY 1509 tcttctgctgagtgcttctctacaacatcaagcccgccagacaaggttgaaacgt 1568
Db 1601 tctgattctgagtgctctctaccacataaagatggogtccaccgggtcgcccaagt 1660
QY 1569 caacaagatcacaaagcggaaattctctgaacggttctgaagatc---ctgaaagaaca 1625
Db 1661 agatatggacatcaagctgaccggacagtcttcctggagcaacacactgctgttccggag 1720
QY 1626 ctgtacgtttgaaatgtggacaacgttgtgacctgtgcccacagagaaggtctgt 1685
Db 1721 catcccccagccagatgagaagtggtgctcactctgagcctgtgaaagagctgagac 1780
QY 1686 catggtgaacggcgtgcgaatcgacaagcctactcgcctccgacgctacagatcat 1745
Db 1781 atatgtaatgggaagctgtgacgggagcgtggtgctgaagtcagggaataggattct 1840
QY 1746 cctgggggatttccacatttttcgattcaaccatcccgaaagaagctcgtgcggaacgca 1805
Db 1841 gatgggcaagaacacagcttttccgcttcaaccacccggagcaggcgtcggaacgga 1900
QY 1806 a 1806
Db 1901 a 1901

RESULT 2
US-09-467-946-2
; Sequence 2, Application US/09467946
; Patent No. 6248594
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: KINESIN-LIKE MOTOR PROTEIN
; FILE REFERENCE: PF-0593 US
; CURRENT APPLICATION NUMBER: US/09/467,946
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: 09/162,373
; EARLIER FILING DATE: 1998-09-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1281811
US-09-467-946-2
```

Query Match 16.2%; Score 380.6; DB 4; Length 3930;
Best Local Similarity 55.0%; Pred. No. 2.6e-103;
Matches 1023; Conservative 0; Mismatches 739; Indels 99; Gaps 10;

```
QY 6 gggcggtggaaatatcaagtggtggtcggttacgccgttcaaccccgagaaatcga 65
Db 80 ggcctggtcctcgtgaaagtggcagtgagggttcggcccttcaaccccgagaaatcga 139
QY 66 ccgtggcgcaaaatattgttgcggatggaaggaaatcaaacatctctccctcc 125
Db 140 ccaggatgccaaagtgtgtggtcagcatgcaggggcaacacccctccatcatcct-- 197
QY 126 ggggtccgaagagaagcgcgtgtaaaagtgcacaaactattatggtgcccgaaggcatt 185
```



```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/428,414A
; FILING DATE: 21-APR-1995
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Kadlecck, Ann T.
; REGISTRATION NUMBER: 39,244
; REFERENCE/DOCKET NUMBER: 210121.407
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANDBERRY
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; OS-08-428-414A-4

```

	Query Match	7.6%	Score 177.8;	DB 2;	Length 3319;
	Best Local Similarity	52.6%;	Pred. No. 7.3e-43;		
	Matches 525;	Conservative 0;	Mismatches 417;	Indels 57;	Gaps 4;
Qy	214	gacaagaatgctcccaactatgcgagacagagaagacactattccaaagatctctcgaggtcccg	273		
Db	710	GACGGGTGCGCGCGACCCCGCGACGAGCAGCATGTGTCGGAGCATCGGTATACC	769		
Qy	274	ctcttgataatgaticcaaggggttaaaacaattgctactcttcgctacggtccagaccggt	333		
Db	770	CTGCTGCAGCAGCGGTCGACGGGTTCAACTCGTGTGTTTGGCTACGGGCACACAGG	829		
Qy	334	tcgggcaagtctctattcaatgatgggcta-----tggcaaggagcat	375		
Db	830	AGCGGGAAGACGTACACGATGATGGCGCGGACGTAGCGCGCTAGTGTGAGGGCAAC	889		
Qy	376	ggcgtgatcccgcgatttgcacagacatgttcgcggcgattataatgaactcagaaagac	435		
Db	890	GCGGTGACCGCGCGGATCTGCTGGAGATCTTTCGGCGGAAGCGCGTGGAGCGCGAG	949		
Qy	436	aagaacctcacttgcacgcgtgaagtttcgtacttggaaatttaacaatgaacgagtgcca	495		
Db	950	GGGCACCTCGCGGTGGATGCTGGACTGGGGTACGTGGAGGTGTACACGACCGCGGTGTCG	1009		
Qy	496	gacttgctgaatccgctcgacaaaggg-----gaattctcaag	531		
Db	1010	GACTGTGTTGGGAGCGGAACAGGGTGTGAAGGCGCGCGAGGAGGTCTACGTGGAC	1069		
Qy	532	gtcgcgaagaacaccgctcgaccggcccttaactgagagacttggcgaagtgtcgtgcga	591		
Db	1070	GTGCGCGAGCACCGAGCGCGCGGCTGTCTCTGGAGGGCAGCGCGCTGTTGGAGTTGGG	1129		
Qy	592	tcattccaagaaatcgaanaattctcatgtagtgagggcaacaaagccagaacggttgcgcgc	651		
Db	1130	AGCCTGGACGATGTGTGCGGCTGTATCGAGATCGGCACACGCGTGCGGCACACCGCTTCG	1189		
Qy	652	acaaatgaacgagacatccagtcgactccacgcgctcttcactttgacctt-----	704		
Db	1190	ACGAAGATGAACGACCGGAGCAGCGCGAGCCACCGGATCATATGCTGCTCTCTGGCGAG	1249		
Qy	705	-----gacgcgaagaagtggcatgtagagagaccacaaatggcacagagaaggttgcgaag	759		
Db	1250	GAGCGGACGATGACGACGAAGACGGGGGAGACGATCCGTACTGCGCGCAAGACGACCGC	1309		
Qy	760	atcagtctggttagatttggcgggtttctgagcgagcaacogtccacgagctacttggagcg	819		
Db	1310	ATGAACCTTGTGGACCTTGGGGGTCTGAGCCGCTGGCGCACTGCAGCTGGAGGGCGAG	1369		

Qy 820 cgactgaagagggtgcagagatcaacccgctcacttltcgacccttagtgctgtgatgcga 879

Db 1370 CAGTTCGAAGGCGCAGCGACATCAACTGTTCGCCTGACGACGCTCGGGCGGTGATCGAC 1429

Qy 880 gcctagcgatatgtcgc---ggaaaaacaagaagaatacagttagtaccttaccga 936

Db 1430 GTGCTCGGACATGCGACGAGGGTGCGAAGGCCGACAGTACAGCGTTGCGCCGTTCCGC 1489

Qy 937 gattcggtactgacgtgggtttctgaaggactccttbggagaaactcgatgaccgccatg 996

Db 1490 GACTCGAAGCTGACGTTTCATCTCTGAAGGACTCGCTTGGCGGGAACTCGAAGACGTTTCATG 1549

Qy 997 atlgccgcatttcgcctgctgaatactaattgaagagactctcagtaaccttcgatat 1056

Db 1550 ATCGCGNCTGTGAGCCCGGCGCTGAACTACGAGGAGAGACGCTGAGCACCGCTCGGTAC 1609

Qy 1057 gcggactctgcgaagcgaatcaagaaccacgcagctgggtcaatgaagacccccgaagcgcgg 1116

Db 1610 CGCTCGCGCGCGCGACATTGTGAATGTTGCGCAGGTGAACGAGGACCCTCGGCGCACGG 1669

Qy 1117 atatccgcgagttgaggaggaactcgcgcagctgaag 1155

Db 1670 CGGATCGCGAGCTGGAGGAGCAGATGGAGGACATGCGG 1708

RESULT 8

PCT-US94-00324-2

; Sequence 2, Application PC/TUS9400324

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven

; TITLE OF INVENTION: Diagnosis of Leishmaniasis

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Immunex Corporation

; STREET: 51 University Street

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98101

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: Apple Macintosh

; OPERATING SYSTEM: Apple System 7.1

; SOFTWARE: Microsoft Word, version 5.1a

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/00324

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/006,676

; FILING DATE: 15-JAN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Perkins, Patricia Anne

; REGISTRATION NUMBER: 34,693

; REFERENCE/DOCKET NUMBER: 5004-WO

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 587-0430

; TELEFAX: (206) 233-0644

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3319 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; ORIGINAL SOURCE:

; ORGANISM: Leishmania chagasi

; PCT-US94-00324-2

Query Match	7.6%	Score 177.8;	DB 5;	Length 3319;
Best Local Similarity	52.6%	Pred. No. 7.3e-43;		
Matches 525; Conservative	0;	Mismatches 417;	Indels 57;	Gaps 4;

Qy 829 gaagggtgcagagatcaaccgctcaatttcgaaccttagtclgtgttgatggcagcgtagcg 888
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 805 gaaggagccaatatcaacaagtcccttgtgactctaggaaattgtcatctccaccttagcc 864
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 889 gatatgc-----gtcgggaaaacagaagaatacagttagtaaccttac 933
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 865 cagaactcccgaagttttcagcagctgccagagctccaacagctcatcttatccccatc 924
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 934 cgagattcgggtactgacgtggctctctgaagagactccttgggaggcaactcgtgatgccgcc 993
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 925 cgagactctgtgtgactcggctcgtcgtgaaggacagccttggaggcaactctaaaaaccatc 984
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 994 atgattgcgcgaatttcgcgctcgtgatattaactttgaagagactctcagtaaccttcga 1053
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 985 attggtgccacgggtgtcctctgcacacactagctacagtgcacatgagcacactgaga 1044
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1054 tatgcggagactctcgsaagcgatcaagaaacacgcagctggttcattgaagaccgcgaacgcg 1113
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1045 tatgcatccagtgccaaaacattatcaaacagccacgcagtaaattgaggatgcgaactta 1104
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1114 cggataatccgcgaggttgaagagaaactcgcgcagctgaggagcaaaactccagagcagt 1173
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1105 aaactgattaggaaactcagagagagattgaaagactgaaagccctgctgtgagcttt 1164
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1174 g 1174
|
Db 1165 g 1165

RESULT 10
US-09-718-841-3
; Sequence 3, Application US/09718841
; Patent No. 6333184
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 633184el motor proteins and methods for
; their use
; FILE REFERENCE: 1051
; CURRENT APPLICATION NUMBER: US/09/718,841
; CURRENT FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1070
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (48)...(111)
; OTHER INFORMATION: n = a, t, c, or g
US-09-718-841-3

Query Match	6.7%;	Score 158.2;	DB 4;	Length 1070;
Best Local Similarity	52.7%;	Pred. No. 2.8e-37;		
Matches 483;	Conservative 0;	Mismatches 398;	Indels 36;	Gaps 5;
Qy 178	aagcatttgcgtgcgtctattggtccctttacaaagaatgctcccaactatgcg	237		
Db 157	atgacatttgcgttattactgtactatggtcgaatccaccagaggatccccagtatgca	216		
Qy 238	agacaggaagaccattatccaagatctcggagtcgccgtctcttggaataatgcattcaaaaggt	297		
Db 217	tctcaagatgggtttccagagtttaggtggaagtactgtctggagttgcgaaggc	276		
Qy 298	tatacaattatattctgcctacggtccagaccggttcggcgaagtctattcaatgatg	357		
Db 277	tatacatatgccttttgcctataggacacagggctctgggaagacatacaccatgctg	336		
Qy 358	ggctatggcaaggagcatggcgatgcctccgcggattttccaggacatgtttccgcggtatt	417		
Db 337	gggacccacgctctgttggtattatcacacacggaatatgtgaaggtctcttcgtcaggaag	396		

RESULT 11
US-09-302-812-38/c
; Sequence 38, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYDROLASES
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES I
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 38
; LENGTH: 29793
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; FEATURE:

RESULT 15
US-09-572-191-1
; Sequence 1, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572,191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4757
; TYPE: DNA
; ORGANISM: Human
US-09-572-191-1

Search completed: July 25, 2002, 05:38:29
Job time: 6617 sec

Query Match 4.5%; Score 106.4; DB 4; Length 4757;
Best Local Similarity 51.1%; Pred. No. 1.8e-21;
Matches 338; Conservative 0; Mismatches 306; Indels 18; Gaps 3;

Qy	462	ttcgtacttgaaattacaatgaacagcgtgcgagacttgcgaatccgtcgacaaagg	521
Db	492	ttcctttattgaattcaacacagcagcatatgatctactgga--ctctgcacgcgc	548
Qy	522	gaatcacaaggtccgagaacaccgctgacccgcccctacgtgaaggacttggcgaagct	581
Db	549	tggactgtacttaaggggagcatatcaagaaggagctcttctgttggcggtggagca	608
Qy	582	ggtcgtgcgacatcccaagaatactcgaatactcatgagtgagggcaacaaagccagac	641
Db	609	ggtgtaacctcagctgtcgaagctatcaggtgttctgtgagatggaggaatagacg	668
Qy	642	ggttgcgcgaacaacatgaacgagacatccagtcgactccacgcgcgtcttcaatttgc	701
Db	669	tgtggcatcaacatcaatgaacagagatcgtctaggtctcatgctcttcaaatc	728
Qy	702	cttgacgcaaaagtggcatgatgaagaccaaataatgacacagagaaggttgcgaagat	761
Db	729	aatagagtcaatggagaaaagtaatgag-----attgtgaatacaggacctccctact	782
Qy	762	cagttcgttagatttggcgggttctgagcgagcaacgtccaccggagctactgagcgcg	821
Db	783	caacctggtgatttagcaggatctgaaaggcaaaagatacccatgcagaaggatgag	842
Qy	822	actgaaggagggtgcagagatcaaccgtcactttcgaccctaggtcgtgtgattgcagc	881
Db	843	attgaaggaaagcaggttaacataaatcgtcattgagctgctgggccaagtattacagc	902
Qy	882	gctagcggtatgtcgtcgggaaacacagaagaataatcagttagcttaccgagattc	941
Db	903	actgtcgacgtgggtaattggaaacacagagaca-----tgtttgctacagagactc	953
Qy	942	ggtactgacgtggcttctgaaggactccttggggagggaactcgatgacccgatgttc	1001
Db	954	caaatcactcttactacggtgattcccttggaggtaatgccaaacacagccataatgc	1013
Qy	1002	cgccattcgcctcgtgatatttaacttgaagagactcagtagctacccttcgatcgga	1061
Db	1014	aaatgttcactcctggatcccggtgttttggggaaacccctatcaacacttaactttgctc	1073
Qy	1062	ctctcgaagcgaatcaagaaccacgcagtggtcaatgaagaccccgacgcgcggatgat	1121
Db	1074	aagagccaagtgtataaacaagcagtagtagtaaatgaagacaccccaaggaaatgtgag	1135
Qy	1122	cc 1123	
Db	1134	cc 1135	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 03:46:42 ; Search time 2969.55 Seconds
(without alignments)
10690.117 Million cell updates/sec

Title: US-09-235-416-2
Perfect score: 2352
Sequence: 1 atgtcggcggtggaatat.....aaacagcgaagcgaattc 2352

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estinu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	321	13.6	1683	11	BC005977
2	277.4	11.8	1074	10	BI906040
3	277.2	11.8	670	9	AI816019
4	270.2	11.5	1263	11	AK020905
5	264.4	11.2	694	10	BI281667
6	260	11.1	761	10	BI414465
7	256.8	10.9	1014	10	BM470467
8	256.4	10.9	770	9	AI363308
9	255.6	10.9	853	9	AU169473
10	255.2	10.9	774	10	BG473130
11	236.8	10.1	670	9	AI879422
12	233.4	9.9	592	10	BG297948
13	233.2	9.9	564	9	AI570691
14	226.4	9.6	586	10	BE252943
15	225.8	9.6	670	10	BF468073
16	224.6	9.5	481	9	AI948449
17	223.4	9.5	690	10	BI685690

C 18	222.4	9.5	507	9	AI339237	AI339237 qt06g12.x
C 19	220.6	9.4	634	9	BE164569	BE164569 RC3-HT047
C 20	220.6	9.4	932	9	AI323149	AI323149 mk16h12.y
C 21	220	9.4	1030	10	BM451369	BM451369 AGENCOURT
C 22	216	9.2	588	9	AV594757	AV594757 AV594757
C 23	215.8	9.2	479	9	AI360405	AI360405 qy85h09.x
C 24	213.2	9.1	561	10	BF150140	BF150140 uy82c07.y
C 25	210.4	8.9	458	10	BE950986	BE950986 UI-M-BH4
C 26	210	8.9	570	10	BF921560	BF921560 MRL-NT017
C 27	208.4	8.9	456	9	AW125600	AW125600 UI-M-BH2
C 28	207.6	8.8	588	9	AW898094	AW898094 RC3-NN006
C 29	206.6	8.8	646	9	AA097269	AA097269 mk16h12.f
C 30	204	8.7	972	10	BM475082	BM475082 AGENCOURT
C 31	202.8	8.6	423	9	AI323139	AI323139 mk10e06.y
C 32	200	8.5	532	9	AU060260	AU060260 AU060260
C 33	198.4	8.4	485	10	BE861354	BE861354 UI-M-AM0-
C 34	198	8.4	660	10	BE258249	BE258249 601115912
C 35	197.2	8.4	457	10	BF548635	BF548635 UI-R-A0-a
C 36	195	8.3	494	9	AI933167	AI933167 wp64d10.x
C 37	193	8.2	799	10	BG393452	BG393452 602411816
C 38	192	8.2	869	10	BE380022	BE380022 601159336
C 39	190.2	8.1	443	9	AA097858	AA097858 mk10e06.f
C 40	189.8	8.1	489	10	BF281801	BF281801 EST446392
C 41	189	8.0	560	10	BG202093	BG202093 602822811
C 42	186.2	7.9	424	9	AA456792	AA456792 zw27h02.f
C 43	185.8	7.9	585	10	BE887181	BE887181 601508384
C 44	181.2	7.7	545	9	AW360994	AW360994 PML-CT024
C 45	177.8	7.6	417	9	BB733392	BB733392 BB733392

ALIGNMENTS

RESULT	1
BC005977	
LOCUS	BC005977 1683 bp mRNA linear HTC 12-JUL-2001
DEFINITION	Homo sapiens, Similar to kinesin family member 13B, clone IMAGE:4103715, mRNA.
ACCESSION	BC005977
VERSION	BC005977.1 GI:14710838
KEYWORDS	HTC.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	1 (bases 1 to 1683)
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. Strausberg, R.
JOURNAL	Direct Submission
TITLE	Submitted (02-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgabbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: CLONTECH Laboratories, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www-shgc.stanford.edu Contact: (Dickson, Mark) mcd@paxil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAL Plate: 21 Row: d Column: 11 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8896163 This clone has the following problem: frame shifted. Location/Qualifiers 1..1683

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone="IMAGE:4103715"  
/tissue_type="Brain, glioblastoma"  
/clone_lib="NIH_MGC_57"  
/lab_host="DH10B"  
/note="vector: pDNR-LIB"
```

BASE COUNT	525 a	345 c	423 g	390 t
ORIGIN				

Query Match	13.6%;	Score 321;	DB 11;	Length 1683;
Best Local Similarity	58.1%;	Pred. NO. 6.5e-75;		
Matches 674; Conservative	0;	Mismatches 460;	Indels 27	

5

QY	1	atgtcgggcggtgaaata	tcaagggtgtgtgtcgggtacgcgcgttcaacgcccgaag	60
Db	19	AGGATGGGGACTC	CAAAGTGAAGTGGCGTTCGGATACGACCCATGAACCGCGGAG	78
QY	61	atcgaccgtggcga	aaattattgtgcggtggaagaaatcaaaccattcctcaacct	120
Db	79	ACTGACTTGCAT	CAAAATGTGTGGTGTGGATGCAACAAGGTTATTTCTTAATCT	138
QY	121	cctcgggtgccga	agaaggcgcgtaaaagtggaactattatgatgcccgaag	180
Db	139	-----GTA	AATACGAATCTTCCAAGGAGATCCCGGGCCACGCCAAG	183
QY	181	gcatttcggttcga	tgcgttatgttcctttgacaagaatg---ctcccaactatgcg	237
Db	184	GTGTTCCTTAT	GCATATCTTTCTGTGCTATGGATGAATCTGTCAAGAAAGTATGCA	243
QY	238	agacagagaac	cattccaaatctcgaagtcgcgcttcgtgataatgattcattcaagggt	297
Db	244	GGTCAGATAT	TGTTTCAAGTGCCCTTGGAGAGAAATATCCTCGAGAAATGCTTTTGATGTC	303
QY	298	tatacaattgtat	cttcgctacgctcagacgcggttcgggcaagtccctattcaatgatg	357
Db	304	TACAATGCAT	GTATCTTGGCTATGCACAGACTGGCTCTGGAAAAATCTTATACCATGATG	363
QY	358	ggctatggcaag	gagcatggcggtatcccgcggtattgcgaagacatgtcccgcggtatt	417
Db	364	GGCAGACTGAC	CAACCTGGATTAATCCCAAGACTTTGTCAGTGGACTCTTTTGAAGCAACT	423
QY	418	aatgaactgcaga	agacaagaccctcacttgcacgctcgaagtttcgtacttgaaatt	477
Db	424	CAGNAGAGGA	AAATG---NAGAACAGAGTTTAAAGTAGAGTGTCTCTACATGGAAATT	480
QY	478	tacaatgaacga	gtgcgagacttgtgaatccgctcgacaaaaggg---gaatctcaaggtc	534
Db	481	TATAATGAA	AAAGTTGCAGACCTCTTGTATCCCAAGGAAGCGCTCAGACGTTGAAAGTC	540
QY	535	cgagaacaccgc	tcgaccgcccctacgtggaggaacttggcgaagctggtcgtgcgata	594
Db	541	AGAGAGCAT	AGTGTGTGGACCTTATGTGCAGCGGACTTTCTTAAACTGGCTGTCAACG	600
QY	595	ttccagaataat	cgaatactcatggtatgaggcacaagaagccagaacggttcgccacca	654
Db	601	TACAAGGAT	ATTGAGTCGTGTGATCTGTGAGGGTTACAAATCTCGCACAGTGTCTGCAACC	660
QY	655	aacatgaacga	gagacatccagtcgatccccacgcgtcttcaactttgaccttgacgcgaaaag	714
Db	661	AACATGAAC	GAGGAGTAGTACCGCATCCCATGCAGTTTTCAAAATCACCTCACATACT	720
QY	715	tggcatgatga	agaagacaaaattggacacagagaagttgcgaagatacgtctgtagat	774
Db	721	CTTACGAT	TGTAAGTCTGGGCATCTGAGAGAAAGTGGGCAAACTCAGCCTGTGTGAT	780
QY	775	ttgcccgggttc	gacgcgaacgctccaccgcgagctacttggacgcgcactgaagaggt	834
Db	781	TTAGCTGGC	AGTGAACGAGCAACGAAGACAGCGCTGCAAGGGACAGGCTGAGGAAGGG	840
QY	835	gcagagataca	aacgcgtcaacttcgacccctaggtcgtgtgattgcgcgctagcgcatg	894

Db	841	AGCAACATTACAAGTCCTCCACAAACCCTCGGTCGTGGTATTCTCAGCTCTTTGCAGATCAG	900
Qy	895	tcgtcggaagaaacagagaagaatcagttagcttaccgagatcggttactgacgtgg	954
Db	901	AGTCTGGCANA--AACAGAATAAATTTGTTCCATATCGTGACTAGTCTCAGTTGG	957
Qy	955	ctttcgaaggactccttggggaggaactcgcgtaccgcgcattgattgcgcgaatttcgct	1014
Db	958	CTGCTCAAAGACACAGCCTCGGGGGTAACAGCAAGACCGCATGGTGGCTACTGTGAGTCCT	1017
Qy	1015	gctgatattaactttgaagagacictcagtcacccttcgatatgcgagctctgcgaagcga	1074
Db	1018	GCAGCTGATACTATGATGAACCCCTTCAACTCTGCGGTATGCAGATPCGAGCCAAAGCAC	1077
Qy	1075	atcaagaaccacgcagtggtgctaagaagaccgcgcgagatgattccgcgagttgaa	1134
Db	1078	ATTGTAAACCACGCTGTGGTGAATGAGGACCCTTAATGCCGAATATCCGGGATCTCCGG	1137
Qy	1135	gaggaactcgcgcagctgagg	1155
Db	1138	GAGAAGTTGAGAACTCCGG	1158

RESULT

2

BI906040/c

LOCUS

DEFINITION

BI906040 1074 bp mRNA linear EST 16-OCT-2001 60306238F1 NIH_MGC_118 Homo sapiens cdna clone IMAGE:5211596 5', mRNA sequence.

ACCESSION

BI906040

VERSION

BI906040.1

KEYWORDS

GI:16168688

SOURCE

EST.

ORGANISM

human.

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM11531 row: c column: 21
High quality sequence stop: 863.

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

high quality sequence stop: 803.
Location/qualifiers

1. .1074

L. :10/4
/organization="Home canion"

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5211596"

```
/clone_lib="NIH_MGC_118"
```

```
/tissue_type="leukocyte"
```

```
/lab host="DH10B"
```

```
/note="Vector: nCMV-spor
```

/home=vector: pcmv-sport/destroyed\ bna source

(destroyed); RNA source

non-activated adult donor

and directionally cloned

cloning). Average inser

1.2-3.3 kb. Library is

full-length clones and w

(Invitrogen) Research

(Inulinogen). Research
this is a NTH MCC ribran

this is a NIH_MGC Librar

a	260 c	220 g	32

11.88; Score 277.4

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

7. ON: MAY 1967 [33]

Matches	524; Conservative	0; Mismatches	326; Indels	11; Gaps	4;
Qy	311	tcttcgctcaggtccagacggttcggtcggaagtcctattcaatgatgggttatggcaag	370		
Db	885	TTTTTGGCTAAGGCCAGACATGGTCTCGGGAAATCTTAATACATCATGGTAAACAAG	826		
Qy	371	-----agcatggtgatcccccggttattgcccagggacatgttccggttatttaataac	424		
Db	825	AACCCCAAGGCTGGCATCCATGCCACATTTATGTGGAAGAACCTTTTAGAGAAAAATCAATG	766		
Qy	425	tgcaagaggacaagaacctcactgcaccgtccgaagtttcgtacttggaatttacaatg	484		
Db	765	ACCACTGATGAGAATGTCTTACTCTGTAGAGGTGACTACATGG-ATTTTACTTGTG	707		
Qy	485	aacgagtcgagactgtcgaatccgtcgcaaaagggaatctcaaggtcccgagaacacc	544		
Db	706	AAAGAGTACGAGATTCTGTAATCCAAAAC-AAGGCTAATTTGCGTGTGCGTGAACACC	648		
Qy	545	cgtagcggccctacgttggagagacttggcgaagcttggctgctgcatatctcaagaaa	604		
Db	647	CAC'TCTTGGACCCATATGTGGAGGATGTG'CCAAAGTTGGCAGTTTACTTCTACACAGCA	588		
Qy	605	tcgaataatctcatgatgagggcaacaaacccagacggttcgcccacaaacatgaacg	664		
Db	587	TTGCTGACCTCATGGATGTGGGAACAACAGGACAGTGGCAGCTACAAACATGAATG	528		
Qy	665	agacatcagtcgatccacgcgtctctcacttgcacttgacgcgaagtggtcatgatg	724		
Db	527	AAACAAGTAGCGGTTCCACGCTGTGTTTACGATTGTTTTCACCCACAGAACACCATTA	468		
Qy	725	aagagacaaaatggacacagagaaggttgcgaagatcagtcctggtatgttggcgggtt	784		
Db	467	ATGAGACCAACCTTTCCACTCAGAAAGTCACTAAATCAGCTTGGTGGATCTAGCAGAA	408		
Qy	785	ctagcgagacaactccacggagctactagcgagcactgaagaggggtgcagagatca	844		
Db	407	GTGAACGAGCTGATTCAACTGGTGCCAAAGGGACTCGATTAAAGGAAGGACCAATATTA	348		
Qy	845	accgctcacttcgacctaggtgtgtgtatgtagcgctagcggtatgtcgtcgggaa	904		
Db	347	ATAAGTCTCTTACAACTTTGGGCAAGTCATTTACAGCTTGGCCGA--GGTGAGTAAA	291		
Qy	905	aacgagagaatacagtttagtacctaccagagatcgttactgacgttggcttctgaag	964		
Db	290	AGAAGAAGAAACAGATTTATTTCCCTACAGGGATTCTGTACTTACCTTCTTCGAG	231		
Qy	965	actccttgaggagcaactcgatgaccccatgattgcgcatttcgctcgtgatatta	1024		
Db	230	ANAATTTAGTGGCAATTCYCGGACTGCAATGGTGTGCTCTGAGCCCGCGGATATCA	171		
Qy	1025	actttgaagagactctcagtcaccttcgatgtatgtagggactctgcgaagcaatcaagacc	1084		
Db	170	ACTACGATGAGACTTTGAGCACTCTGAGATATCGATATCGTGCACAAACAAATTAATGCA	111		
Qy	1085	acgcagtggttcaatgaagaccgcaacgcgcggatgatccccgagttgaagaggagaaactcg	1144		
Db	110	ATGCTGTATCAATGAGGACCCCAATGCCAACTGGTTCGTGTAATTAAGAGGAGGCTGA	51		
Qy	1145	cgcagctgagagacaaactcc	1165		
Db	50	CACGGCTGAAGGACCTTCTTTC	30		
RESULT	3				
LOCUS	AI1816019/c	AI1816019	670 bp	mRNA	linear
DEFINITION	au44c12.x1 Schneider fetal brain 00004 Homo sapiens cDNA clone IMAGE:2517622 3' similar to SW:KF1A_HUMAN Q12756 KINESIN-LIKE PROTEIN KIF1A ;, mRNA sequence.				
ACCESSION	AI1816019				
VERSION	AI1816019.1	GI:5431565			
KEYWORDS	EST.				
SOURCE	human.				

Db	649	GGACTCGATTANAGGAAGGAGCAAAATATTATAAGTCTCTTACAACTTTGGGCAAAAGTCA	708
QY	875	ttgca-----gcgctagcgggatatcgctcgggaaaaacagaagaatacag-----	921
Db	709	TTTCAGCCTTNGGCCGAGTGGATAACTGCACCTAGNCAGACTAAAAGAGAAGAAACC	768
QY	922	----ttagtaaccttaccagagattggctactgacgggactctctgaaggactccttggagag	977
Db	769	GATTTTATTTCCCTACAGNCGATCTGTACTTACTTGGCTCCTTCGAGAAAATTTTANGTGG	828
QY	978	caactcgatgaccgccatgatgccgcatttgccttgcctgctgataacttaactttgaagac	1037
Db	829	CAATTCTCGNACTGCCAATGGTGTCTGTGAGCCCGCGGATATCCACTACGATGAGAC	888
QY	1038	tctcagtaacctt-cgatatcggaactctggaagcgaatcaagaacca	1085
Db	889	TTTGAGCAGCTCTTGAAATATGCAGATCGTGCAAAACCAAAATAAATGCA	937
RESULT	8		
ALIAS	AI363308/c		
LOCUS	AI363308	770 bp	mRNA
DEFINITION	qy57c04.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2016102 3' similar to SW:KFLA_HUMAN Q12756 KINESIN-LIKE PROTEIN KIF1A ;, mRNA sequence.		
ACCESSION	AI363308		
VERSION	AI363308.1	GI:4114929	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	NCI/WINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
TITLE	National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index		
JOURNAL	Unpublished (1998)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html Insert Length: 1090 Std Error: 0.00 Seq primer: -400P from Gibco High quality sequence stop: 407.		
FEATURES	Location/Qualifiers		
source	1..770 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2016102" /clone_lib="NCI_CGAP_Brn23" /tissue_type="glioblastoma (pooled)" /lab_host="DH108" /note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCACTCTCAAGTGGGCGCGGCATCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."		
BASE COUNT	125 a	233 g	199 t
ORIGIN	2 others		


```
Query Match      10.9%; Score 256.4; DB 9; Length 770;
Best Local Similarity 65.3%; Pred. No. 9.5e-58;
Matches 409; Conservative 0; Mismatches 208; Indels 9; Gaps 2;

QY 284 atgcttcaagggttatcaaatgtatcttcgctacggtcagaccggttcggcaagt 343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 629 ATGCCCTTGAGGATACACCGTGTGCTATTTGGCTATGGGAGACGGTTCGGGCAAGT 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 344 ctattcaatgatgggttatgcaaggagcat-----ggcgtgctccgggattgcc 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 569 CNTACACAATGATGGCAAGCAGGAGAGGACGACGAGGGGCATCATCCACAGCTCTGG 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 398 agacacatgttcggcggtattaatgaactgcagaagacaaacactcaactgcacgtcg 457
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 509 AGGACCTCTCTCGGGATCAACGACACGACGACCAACGAC---ACANTGCTCTACTCCGTG 453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 458 aagtttcgtacttgaaattcaaatgaacgagtgctgagactgtgtaactcgtgcagaa 517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 452 AGGTCAGCTACATGGAGATTACTGTGAGCGGCTCGTGACCTCTGAACCCCAAGAACA 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 518 aggggaatctcaaggtccgagaacacccgtcgaccggcccttaactgagacttgccga 577
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 392 AGGGCAACCTTCGCGTGAGGGAGCACCCACACTGCTGGGGCCCTACGTGAGGACCTCTCCA 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 578 aactgtgtcgatcattccaagaatacgaataatctcatggtgagggcacaagaagcca 637
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 332 AGCTGGCTGTCACTTCTTACATGACATGACATGACGAGGACCTCATGGACTCAGGGCAAGGCCA 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 638 gaacggttgcgcgcacaaacatgaacgagacatccagtcgatccacacgctcttcaatt 697
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 272 GGACGCTGGCGGCCACCAACATGATGAGACGACGAGTCTGCTCCACGCCCTCTCAACA 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 698 tgacattgacgcacaaagtggcgtatgtagagagaccacaaatggacagagaggttcga 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 212 TCATCTTCCACCAGAGAGCGCATGACGAGAGACCAATATCACCACGAGGAAGGTGAGCA 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 758 agatcagctctgtagatttgcgggttctgagcagacacgtccacaggaactactggag 817
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 152 AAATCAGCTGGTGGACCTGGCTGGGACGAGCGGGCTGACTCCACGGAGGCCAAGGCCA 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 818 cgcgactgaaggaggtgcagagatcaaccgctcaacttcgaccctaggtcgtgattg 877
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 92 CGCGCTCAAGAGGGGGGCCAACATCAACAAGTCGCTGACCACCTCGGGCAAGGTCTACT 33
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 878 cagcgctagcggtatgtcgtcgga 903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 32 CGGCCTGGCTGAATGGACTCCGGA 7
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
LOCUS AU169473/3
DEFINITION AU169473 Ol-br-ad cDNA Oryzias latipes cDNA clone br4665, mRNA
sequence.
ACCESSION AU169473
VERSION AU169473.1 GI:12591542
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

REFERENCE
1 (bases 1 to 853)
Mita,K., Ishikawa,Y. and Yamauchi,M.
Establishment of cDNA database of medaka, Oryzias latipes
Unpublished (2001)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmits@irs.go.jp
method:uni-directional sequence direction:sequenced from T3 primer
```

```
FEATURES             (5' -> 3').
source               1..853
                    /organism="Oryzias latipes"
                    /strain="HNI"
                    /db_xref="taxon:8090"
                    /clone="br4665"
                    /clone_lib="Ol-br-ad cDNA"
                    /sex="female/male mixed"
                    /tissue_type="brain"
                    /dev_stage="adult"
BASE COUNT          165 a      210 c      208 g      270 t
ORIGIN
```

```
Query Match      10.9%; Score 255.6; DB 9; Length 853;
Best Local Similarity 62.3%; Pred. No. 1.7e-57;
Matches 440; Conservative 0; Mismatches 254; Indels 12; Gaps 2;

QY 221 atgtcccaactatgcgagacaggaagaccctattccagatctcgagatccggtcttcg 280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 722 AGGATGTCAATTACGCTCTCAGATGCAAGTTTACAAAGATATTGGAGAAATGTTGC 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 281 ataatgcattcaagggtttatacaaatgtatcttcgctacggtcagaccggttcgggca 340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 662 TCCACGCTTTGAAGGATACAACTGCTGCATCTTTGCTACGGTCAGACAGCGCTGGCA 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 341 agtccattcaatgatgggtatggcaagg-----agcatggcggtatcccgga 391
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 602 AGTGTACACATGATGGGGAAGCAGGATGTCAAGGACCAGCAAGGCATCATTCGACTGC 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 392 ttgccaggacatgttccggcggtattaatgaactgcagaagacaaacactcactgca 451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 542 TGTGTGAAGACTGTTTCAGAAAAATCAGCGA---CAACACAGACAACAGCATGCTTACT 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 452 ccgtcgaagttcgtacttggaatttacaatgaacgagtcgagacttgctgaatccgt 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 485 CTATTGAGGTCAGTACTATGGAGATCTACTGTGAACGTGTACGGGACTTGTGAACCCCA 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 512 cgacaaaggggaattctcaaggtccgagacacccgtcgaccgcccctacgtgagact 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 425 AAAACAAAGGCAACCTGCGCGTGGGAGAGCATCTCTGTATGGGGCCCTATGTGGAGACC 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 572 tggcgaagctggtcgtcgatcatccaagaatacgaataatctcatgtagagggcaaca 631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 365 TGTCAAAGCTGGCTGTACCTCGTACAAATGACATCCAGGACCTGATGATTCTGGCAACA 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 632 aagccagaacggttgcgccacacaaacatgaacgagacatccagtcgatccacgctct 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 305 AGGCCAGACTGTGGCTGGCCACCACCAATGAACGAGACGAGCTCGCTCGCTGCTGCT 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 692 tcacttgaccttgacgcaaaagtggcatgatgaagagacacaaatggacacagagaag 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 245 TCAACATCATATTACCTCAGAAAGCCCATGATCGAGAGACCGACACACTCGGAAAGG 186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 752 ttgcgaagatcaactctggttagatttggcggttctgagcgcagacacgtccacggagcta 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 185 TCAGAAAAATCAGCTTGGTGGATTGGCCGGAGGAGGCGGATTCGACTTGGAGCCA 126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 812 ctgagcgcaactgaagaggggtgcagagatcaaccgtcaccttcagacctaggtcgt 871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 AAGGAACGAGCTTAAAGAGGAGCAACATCAACAAATCTCTGACCACCTCTGGGGAAG 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 872 tgattgcagcgtagcggtatgtcgtcggggaaacagaagaaga 917
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 65 TCATCTCTGCTTTGGCAGAAATGGATTCTGGACCAATAAGAATAA 20
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 10
LOCUS BG473130
DEFINITION 602515112F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4646835 5',
mRNA sequence.
```


Job time: 3142 sec

THIS PAGE BLANK (USPTO)